Dissertationes Scholae Doctoralis Ad Sanitatem Investigandam
Universitatis Helsinkiensis No. 20/2018

STUDIES ON THE GENETICS OF HEART FAILURE

Perttu Päiviö Salo

Genomics and Biomarkers Unit,
Department of Public Health Solutions,
National Institute for Health and Welfare,
and
Doctoral Programme in Integrative Life Science,
Institute for Molecular Medicine Finland (FIMM),
Faculty of Medicine, University of Helsinki,
Helsinki, Finland

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Medicine of the University of Helsinki, for public examination in lecture hall 3, Biomedicum 1, on 27.4.2018 at 12 noon.

Helsinki 2018

ISBN 978-951-51-4170-5 (pbk.)
ISBN 978-951-51-4171-2 (PDF)
ISSN 2342-3161 (print)
ISSN 2342-317X (online)

Helsinki University Printing House
Helsinki 2018

Supervised by

Research professor Markus Perola
National Insitute for Health and Welfare,
Department of Public Health Solutions, Genomics and Biomarkers Unit,
Helsinki, Finland

University of Helsinki,
Diabetes and Obesity Research Program,
Helsinki, Finland

Reviewed by

Adjunct Professor Mikko Pietilä
Turku University Hospital,
Heart Center,
Turku, Finland

Professor Arto Mannermaa
University of Eastern Finland,
School of Medicine,
Institute of Clinical Medicine, Pathology and Forensic Medicine,
Kuopio, Finland

Opponent

Professor Katriina Aalto-Setälä
Tampere University Hospital, Heart Center
University of Tampere, Faculty of Medicine and Life Sciences
University of Tampere, BioMediTech,
Tampere, Finland

When you stop at the red light
look at the red light
and smile

-Thich Nhat Hanh

To my family

ABSTRACT

Heart failure is a complex clinical syndrome, most often caused by ischemic, hypertensive, and valvular heart disease. In this thesis, we studied how genetic variation contributes to hypertension and acute coronary syndrome, the most frequent causes of heart failure. We also studied the role of genetic variation in Takotsubo cardiomyopathy, a disease which still remains poorly understood and often presents with transient heart failure.

To identify trait-associated genetic variants, we analyzed several population-based samples of Finns using genetic association tests. Three of the four publications constituting this thesis are genome-wide association studies (GWAS).

In the first study, we identified variants near the *PRDM6* gene, which associated with both higher systolic blood pressure and a higher risk of intracranial aneurysms. This result was one of the first to show a shared genetic background for blood pressure and intracranial aneurysms. Based on data on the function of *PRDM6* and other trait associations in the locus, we hypothesize that the genetic variants at this locus increase systolic blood pressure and aneurysm risk via an effect on the proliferation of smooth muscle cells in the arterial wall.

In the second study, we performed a GWAS separately on the two subtypes of myocardial infarction in acute coronary syndrome, the ST-segment elevation and non-ST-segment elevation myocardial infarctions (STEMI and NSTEMI). Genetic variation near *DRAM2*, encoding a protein participating in the regulation of autophagy, increased the risk of NSTEMI with little to no effect on the risk of STEMI. This finding is rather surprising, given that both infarction types are mostly the result of the same disease process, namely coronary heart disease.

The third study addressed atrial and B-type natriuretic peptides (ANP and BNP), which are secreted by cardiomyocytes, both during normal homeostasis and especially during acute heart failure. We identified a new locus near the calcineurin subunit gamma gene *PPP3CC*, which was

associated with the ratio of the circulating concentrations of active BNP and the N-terminal fragment of the proBNP prohormone. Based on the genetic results, we also showed that the data specifically supports the blood-pressure-lowering effect of ANP in the general population, and not that of BNP.

In the fourth study, we studied the possible genetic background of Takotsubo cardiomyopathy. The cause of this condition is currently unknown but a contribution of genetic variation has been suggested in previous literature. We could not replicate previous genetic findings in the disease to show that if genetic predisposition exists, it is similar to that of complex diseases, such as acute coronary syndromes, with small absolute risks conferred by each of multiple genetic risk variants.

The results of this thesis provide new information on the molecular basis of the two most common causes of heart failure: hypertension and acute coronary syndromes. They also limit plausible genetic models that are compatible with observational data for Takotsubo cardiomyopathy.

TIIVISTELMÄ

Sydämen vajaatoiminta on oireyhtymä, jonka merkittävimmät syyt ovat sepelvaltimotauti, verenpainetauti ja läppäviat. Tässä väitöskirjatyössä tutkittiin perinnöllisen muuntelun yhteyttä sepelvaltimotautiin ja verenpainetautiin, jotka ovat sydämen vajaatoiminnan tärkeimpiä syitä. Lisäksi väitöskirjassa selvitettiin Takotsubo-sydänlihastaudin mahdollista perinnöllistä taustaa. Takotsubo-sydänlihastauti on harvinainen sairaus johon usein liittyy ohimenevä sydämen vajaatoiminta.

Väitöskirjatyö perustuu geneettisille assosiaatiotesteille suomalaisissa väestöaineistoissa. Kolme väitöskirjan neljästä osatyöstä on perimänlaajuisia assosiaatiotutkimuksia (GWAS). Ensimmäisessä osatyössä tunnistimme geenin *PRDM6* lähellä alleeleita, jotka ovat yhteydessä korkeaan verenpaineeseen ja aivovaltimopullistumien esiintymiseen. Nämä tulokset ja *PRDM6*-geenistä aiemmin julkaistut tiedot viittaavat siihen, että *PRDM6*:n geenilekspressio valtimonseinän sileissä lihassoluissa saattaa vaikuttaa sekä valtimoverenpaineeseen että aivovaltimonpullistumariskiin.

Toisessa osatyössä teimme perimänlaajuisen assosiaatiotutkimuksen kahdesta sydäninfarktityypistä, ST-noususydäninfarktista (STEMI) ja sydäninfarktista ilman ST-nousua (NSTEMI). Osoitimme, että geneettinen muuntelu kromosomilla 1 lähellä *DRAM2*-geeniä altistaa NSTEMI-infarktityypille, ja että näiden alleelien vaikutus ST-nousuinfarktiriskiin oli merkittävästi pienempi. Löydös on yllättävä, sillä molemmat infarktityypit ovat sepelvaltimotaudin päätetapahtumia.

Kolmannessa osatyössä tutkimme perinnöllistä muuntelua sydämen erittämien natriureettisten peptidien määrässä ja sen vaikutusta verenpaineeseen. Eteispeptidi (ANP) ja B-tyypin natriureettinen peptidi (BNP) ovat sydänlihaksen verenkiertoon erittämiä hormoneita, joiden eritys lisääntyy sydämen vajaatoiminnan yhteydessä. Suurempaan eteispeptidin määrään liittyvät geenimuodot olivat yhteydessä matalampaan valtimoverenpaineeseen ja verenpainetautiriskiin, mutta BNP-määrään liittyvien geenimuotojen ja verenpaineen välillä ei ollut tilastollisesti

merkitsevää yhteyttä. Tutkimuksessa tunnistettiin myös harvinainen suomalaiseen väestöön rikastunut BNP-geenimuoto, joka saattaa häiritä aminoterminaalisen pro-B-tyypin natriureettisen peptidin (NT-proBNP) käyttöä sydämen vajaatoiminnan erotusdiagnostiikassa.

Neljännen osatyön perimänlaajuinen assosiaatiotutkimus Takotsubo-sydänlihastaudista osoitti, että suomalaisessa väestössä ei todennäköisesti ole tälle taudille voimakkaasti altistavia yleisiä geenimuotoja. Aiemmin julkaistut tulokset taudin mahdollisesta periytyvyydestä eivät toistuneet tutkitussa aineistossa. Tulokset eivät tue kirjallisuudessa esitettyä ajatusta Takotsubo-sydänlihastaudin periytyvyydestä.

Väitöskirjatyön tulokset antavat uutta tietoa sydämen vajaatoiminnan merkittävimpien syiden mekanismeista ja rajaavat Takotsubo-sydänlihastaudin mahdollista perinnöllistä taustaa.

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following publications, referred to using their roman numerals in the text:

I Gaal EI, **Salo P***, Kristiansson K*, Rehnstrom K, Kettunen J, Sarin AP, Niemela M, Jula A, Raitakari OT, Lehtimaki T, Eriksson JG, Widen E, Gunel M, Kurki M, von und Zu Fraunberg M, Jaaskelainen JE, Hernesniemi J, Jarvelin MR, Pouta A, International Consortium for Blood Pressure Genome-Wide Association Studies, Newton-Cheh C, Salomaa V, Palotie A, Perola M. Intracranial aneurysm risk locus 5q23.2 is associated with elevated systolic blood pressure. *PLoS Genet* 2012, 8: e1002563.

II **Salo PP**, Vaara S, Kettunen J, Pirinen M, Sarin AP, Huikuri H, Karhunen PJ, Eskola M, Nikus K, Lokki ML, Ripatti S, Havulinna AS, Salomaa V, Palotie A, Nieminen MS, Sinisalo J, Perola M. Genetic Variants on Chromosome 1p13.3 Are Associated with Non-ST Elevation Myocardial Infarction and the Expression of DRAM2 in the Finnish Population. *PLoS One* 2015, 10: e0140576.

III **Salo PP**, Havulinna AS, Tukiainen T, Raitakari O, Lehtimäki T, Kähönen M, Kettunen J, Männikkö M, Eriksson JG, Jula A, Blankenberg S, Zeller T, Salomaa V, Kristiansson K*, Perola M*. Genome-wide association study of atrial and B-type natriuretic peptides implicates ANP rather than BNP in the regulation of blood pressure in the general population. *Circulation: Cardiovascular Genetics* 2017, accepted for publication.

IV **Salo PP***, Parkkonen O*, Perola M, Paunio T, Lokki ML, Nieminen MS, Sinisalo J. Genome-wide association study of Takotsubo cardiomyopathy. Submitted.

* Equal contribution

Publications I & II have been included previously in the doctoral theses of Emília Ilona Gaál ("On the genetics of intracranial aneurysm and on growth factor induced angiogenesis in the murine brain". University of Helsinki, 2012) and Satu Vaara ("Clinical and genetic risk factors in acute coronary syndromes". University of Helsinki, 2017).

ABBREVIATIONS

ACS	Acute coronary syndrome
ANP	Atrial natriuretic peptide
BNP	B-type natriuretic peptide
bp	base pair(s)
BP	blood pressure
CAD	coronary artery disease
CEPT1	choline/ethanolamine phosphotransferase 1 gene
COROGENE	The COROGENE Study
cTnI	cardiac Troponin I
D'	Lewontin's D-prime
DBP	Diastolic blood pressure
DNA	Deoxyribonucleic acid
DRAM2	DNA damage regulated autophagy modulator 2 gene
ECG	Electrocardiography
EF	Ejection fraction
ESC	The European Society of Cardiology
FINRISK	The National FINRISK Study
GALNT4	polypeptide N-acetylgalactosaminyltransferase 4 gene
GTEx	The Genotype-Tissue Expression project
GWA	Genome-wide association
GWAS	Genome-wide association study
H2000	Health 2000 Survey
HBCS	Helsinki Birth Cohort Study
HDL	High-density lipoprotein
HF	Heart failure
HUCH	Helsinki University Central Hospital
LD	Linkage disequilibrium
LDL	Low-density lipoprotein
LDLR	Low-density lipoprotein receptor

LOF	Loss-of-function
LVEF	Left ventricular ejection fraction
MAF	Minor allele frequency
MI	Myocardial infarction
mmHg	millimeter of mercury
mRNA	messenger RNA
NCBI	National Center for Biotechnology Information
NFBC	Northern Finland Birth Cohort
NPPA	natriuretic peptide A gene
NPPB	natriuretic peptide B gene
NSTEMI	Non-ST elevation myocardial infarction
NT-proBNP	aminoterminal pro-B-type natriuretic peptide
PCSK9	proprotein convertase subtilisin/kexin type 9 gene
PPP3CC	protein phosphatase 3 catalytic subunit gamma gene
r^2	Squared correlation coefficient
RNA	Ribonucleic acid
SBP	Systolic blood pressure
SNP	Single nucleotide polymorphism
STEMI	ST-elevation myocardial infarction
TACOS	Tampere Acute Coronary Syndrome Study
TTC	Takotsubo cardiomyopathy
UA	Unstable angina pectoris
YFS	The Cardiovascular Risk in Young Finns Study

INTRODUCTION

Variation in most human traits is to some degree caused by genetic differences. The identification and mechanistic understanding of specific genetic differences in traits will likely eventually lead to a detailed understanding of how diseases, such as coronary heart disease, hypertension, or Alzheimer's disease, develop. This understanding, in turn, can be used to develop efficient preventive measures and treatments.

Technological and scientific advances in the 21st century have resulted in methods enabling large-scale investigation of human genetic variation. This thesis comprises four studies applying these methods to identify genetic differences in traits relevant to the onset of heart failure. Rather than an independent disease, heart failure is a syndrome that may be caused by various diseases. Most notably, coronary artery disease and hypertensive heart disease are both the main causes of heart failure and also among the leading causes of death world-wide [1,2]. In the vast majority of cases, heart failure is a multifactorial or complex trait, caused by many genetic and environmental factors acting in concert. [3,4]

The studies presented here are based on data from the Finnish population. The characteristics of genetic variation in Finns have favoured genetic investigation, particularly genetic research in monogenic disease focusing on pedigrees [5]. Studies of common complex disease, on the other hand, depend on large, well-characterised study samples. Population-based cohorts of Finns, such as the National FINRISK Study cohorts, are thus central to the work presented in this thesis. However, results from one population are not limited to that population alone. Although differences between human populations do exist, they are rather small and results from one population are mostly applicable to others as well [6,7].

1 REVIEW OF THE LITERATURE

1.1 HEREDITY

Living beings pass on some of their traits to their offspring. This transmission of traits from one generation to the next, called heredity or inheritance, is one of the defining features of life. The scientific study of inheritance was pioneered in the 19th century by Gregor Mendel, an Augustinian monk. Working on the common pea plant, he discovered fundamental mathematical rules describing the distribution of traits, such as flower color in the offspring when crossing different pea plants. Although he was certainly not the first person to study heredity, Gregor Mendel is regarded as the founder of the modern scientific study of inheritance, nowadays known as genetics.

The basic unit of hereditary information is the gene. Genes determine the molecular structure and function of cells. On the molecular level, genes of cellular organisms are composed of deoxyribonucleic acid or DNA [8]. Solving the structure of DNA in 1953 and, subsequently, deciphering how genetic information is stored in DNA, transcribed into RNA, and translated into proteins are among the greatest scientific achievements ever made [9-11]. The understanding of how genes are physically encoded into DNA provides us with a mechanistic explanation of how traits can pass from one generation to the next.

Since Mendel's days, great effort has been made in trying to understand how, exactly, genes influence different traits in various organisms. This is not a trivial task. After all, genes are somehow able to guide the development of living beings, arguably the most complex entities known to science.

1.2 DNA

1.2.1 STRUCTURE OF DNA

DNA is composed of nucleotides. Linked together by covalent bonds, nucleotides contain a nitrogenous base, either adenoside (A), thymine (T), cytosine (C) or guanine (G), which together form the letters of the DNA code. Two chains or strands of DNA are bound together via hydrogen bonds so that C always binds to G on the opposite strand and A always binds to T. When the two strands of a DNA molecule are bound by base-pairing, they curl up together as though wrapping around a common axis, forming the iconic double helical shape of DNA. [12]

1.2.2 MUTATION AND RECOMBINATION

Mutations, errors in maintaining DNA's nucleotide sequence, give rise to genetic variation between individuals of a given species. They are also the reason why the genomes of different species are dissimilar. Mutations occurring in the germline cells can pass from parents to their offspring, while mutations altering the genetic sequence in somatic cells cannot.

Mutations produce new alleles, alternative genetic sequences in a given genomic location or locus. Point mutations are the simplest type of mutations, changing one nucleotide into another. Deletions and insertions either remove or insert one or more nucleotides. Duplication and amplification are special cases of insertions, increasing the number of copies of a given genetic sequence. Translocations delete a stretch of DNA but insert it to a new location in the genome. Inversions flip a sequence of DNA around, otherwise maintaining its position and sequence in the genome.

Variant alleles, the results of mutations, are divided into single-nucleotide polymorphisms (SNPs) produced by point mutations, indels originating from insertions and deletions, and more complex alleles such as translocations and inversions [12]. Some of the complex alleles are often grouped together under the terms copy number variation (CNV) and structural variation (SV) [6,13].

During meiosis, the homologous chromosomes of each chromosome pair align against each other and physically exchange stretches of DNA. This recombination of chromosomes shuffles alleles between homologous chromosomes, producing chromosomes with new combinations of alleles. Chromosomes then segregate randomly so that each mature sex cell receives only one copy of each chromosome. Errors in this process may occur, causing aneuploidies, where the number of chromosomes in the cell is abnormal. Mutation and recombination are the driving forces creating genetic diversity in populations, producing new alleles and creating new combinations of alleles.

1.3 ALLELES IN POPULATIONS

1.3.1 ALLELE AND GENOTYPE FREQUENCIES

The distribution of alleles in populations are described by two central metrics, their allele and genotype frequencies. The frequency of an allele is its relative frequency in a population or a sample, i.e. the proportion of individual chromosomes carrying said allele. In biallelic loci, the rarer allele is referred to as the minor allele. The frequency of the more common allele is thus 1 - the minor allele's frequency (MAF).

Genotype frequencies, or the proportion of genotypes in a sample or population, depend on allele frequecies. The Hardy-Weinberg principle shows that, in the absence of evolutionary forces, genotype frequencies depend *only* on allele frequency. For a biallelic locus with alleles p and q, the genotype frequencies are given by

$$1 = p^2 + 2pq + q^2$$

where p^2, $2pq$, and q^2 are the genotype frequencies for the pp homozygotes, the heterozygotes, and the qq homozygotes, respectively [14].

1.3.2 LINKAGE DISEQUILIBRIUM

When a mutation occurs on a given chromosome, the resulting mutant allele is physically linked to all other alleles already present on said chromosome at other loci. Until the physical link between the alleles is broken by recombination, the alleles are inherited together as a haplotype, introducing correlation between the alleles. As the probability of a recombination event between two chromosomal positions is proportional to the distance between the positions, alleles close to each other are more strongly correlated than alleles separated by longer distances. The distribution of recombination events is not uniform, however, and some genomic regions are much more prone to recombination events ("recombination hotspots") than others [15].

Correlation or non-independence between alleles in different genomic loci is collectively denominated linkage disequilibrium (LD). When two alleles are in linkage equilibrium, the population frequency of chromosomes carrying both alleles is the product of the alleles' frequencies, as given by the definition of independent events in probability theory. Formally, given a locus with alleles A and a with frequencies p_A and q_a, and a second locus with alleles B and b with frequencies p_B and q_b, the frequency P_{AB} of chromosomes carrying alleles A and B is given by the alleles' frequencies and the linkage disequilibrium parameter D [16]:

$$P_{AB} = p_A p_B + D$$

The linkage disequilibrium parameter D depends on allele frequencies, making it somewhat difficult to compare values of D for different pairs of alleles. A more easily interpreted value D' is commonly used, expressing D as the proportion of its greatest (if positive) or smallest (if negative) possible value given the frequencies of the two alleles [17]. Often, the sign is omitted and only the absolute value is reported. This ensures that D' always falls within [-1, 1] or [0, 1], making it more meaningful and easy to compare:

$$D' = \frac{D}{D_{MAX}} = \begin{cases} \frac{D}{min\{q_a p_B,\ p_A q_b\}}, D \geq 0 \\ \frac{D}{min\{q_a q_b,\ p_A p_B\}}, D < 0 \end{cases}$$

It is often natural to interpret linkage disequilibrium in terms of how well knowing the genotypes of one of the alleles can predict the other. D' partly answers this, as it directly describes how often the rarest of the alleles at the two loci is met on the same chromosome with the more common alleles in the other locus. If the absolute value of D' equals one, the rarest of the four alleles perfectly predicts the genotypes at the other locus, being either always (D' = 1) or never (D' = -1) on the same chromosome with one of the two more common alleles at the other locus. The opposite, however, does not apply, as D' does not readily describe how well the more common allele predicts the rarer allele. The squared correlation coefficient r^2 overcomes this asymmetry and provides a general measure of correlation between two alleles [18]:

$$r^2 = \frac{D^2}{p_A q_a p_B q_b}$$

Compared to D', the squared correlation coefficient r^2 captures another side of the relationship between two alleles at two loci. This is particularly evident in how differences in allele frequencies affect the two measures. A large difference in allele frequencies guarantees a small value of r^2, while D' is independent of allele frequency differences. D' is perhaps biologically more meaningful, as a new mutant allele has D' = 1 with all other alleles on the chromosome until recombination breaks the linkage. The squared correlation coefficient r^2 is a more abstract statistical measure of linkage disequilibrium.

1.4 THE HUMAN GENOME

1.4.1 STRUCTURE

Due to genetic variation, the genomes of any two individuals are different, barring the special case of clones such as monozygotic twins. Thus, *the* human genome as a concept is an abstract idea, which does not exactly correspond to the genome of any specific human. In practice, it is used as a shorthand for "a typical human genome" or in reference to a specific model of the genomes of humans[1].

Most humans inherit one copy of each nuclear chromosome from both parents, making humans a diploid species. Oocytes and sperm cells each carry a haploid human genome, containing approximately 3.2 billion base pairs of DNA distributed into 23 nuclear chromosomes (Figure 1, Table 1) [19,20]. Chromosome 1, the largest human chromosome, contains approximately five times as much DNA as the smallest, chromosome 21. Some chromosomes contain few genes relative to their physical size, while e.g. the diminutive chromosome 19 packs approximately 25 genes per million base pairs (Mbp), four times the genomic average of 6.5 genes per Mbp [21].

In addition to the nuclear chromosomes, oocytes and sperm cells also carry mitochondria, which have multiple copies of their own small chromosome. Oocytes contain approximately 10^6 mitochondria, whereas the sperm cells hold few mitochondria, which are destroyed when the sperm enters the oocyte [22]. Thus, in contrast to the nuclear chromosomes, the multiple copies of the mitochondrial chromosome are inherited from the mother only.

As the structure and contents of the human genome are the result of its evolutionary history, sequence conservation can be used to detect functionally constrained elements in the genome. Approximately 5.5% of the entire human genome is under purifying natural selection, while the remaining 94.5% appears to evolve neutrally unaffected by natural selection [23]. Although sequence conservation during evolution is not a perfect measure of biological function, this strongly suggests that the bulk of the

[1] E.g. the Genome Reference Consortium human reference genome assembly GRCh38.p11

DNA sequence in the human genome carries little biologically important information.

Roughly 50% of the human genome consists of repetitive DNA sequences which contain few genes, such as the tandemly repeat sequences present at centromeres and telomeres [19]. Most of the repetitive elements of the genome are formed by inactive transposons. Hundreds of thousands of inactive transposons are found throughout the nuclear genome, together comprising a striking 44% of its total size. Compared to the number of inactive transposons, there are few active transposons, producing an average of 0.1 to 0.01 new insertions per live birth. [24] Furthermore, little if any of the repetitive DNA in the human genome has been under purifying natural selection [23]. The repetitive DNA sequences in the human genome are thus mostly inactive transposable elements and appear to serve no sequence-specific purpose.

The conserved 5.5% of the human genome contains genes and regulatory elements. Proteins are the main drivers of biochemical activity in cells, with the notable exception of protein synthesis itself, but the exons of classical protein-coding genes occupy only 1.5% of the human genome. Most of the functional elements of the human genome must, therefore, serve a regulatory function, making sure the right amount of proteins are expressed in the right cells at the right time.

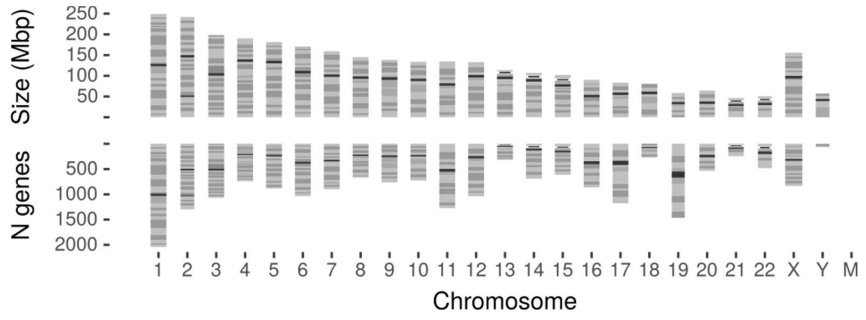

Figure 1. The human genome. The sizes of the human chromosomes are shown in base pairs (top row) as well as relative to the number of protein coding genes on each chromosome (bottom row). The physical size of chromosomes only partially correlates with the amount of genes they contain, due to the uneven genomic distribution of a vast amount of mostly repetitive DNA encoding no genes. Figure based on the GRCh38p11 reference genome and the GENCODE v27 gene annotation set [20,21]. N.B.: The mitochondrial chromosome is too small to be visible at this scale.

Table 1. The human genome.

Chr	Size (bp)*	Protein-coding genes†	Other genes†	Genes per Mbp (protein-coding / other)
total	3,088,269,832	20,036	23,544	6.49 / 7.62
1	248,956,422	2,040	1,988	8.19 / 7.99
2	242,193,529	1,301	1,652	5.37 / 6.82
3	198,295,559	1,072	1,184	5.41 / 5.97
4	190,214,555	747	1,024	3.93 / 5.38
5	181,538,259	882	1,278	4.86 / 7.04
6	170,805,979	1,035	1,020	6.06 / 5.97
7	159,345,973	902	1,091	5.66 / 6.85
8	145,138,636	668	1,083	4.6 / 7.46
9	138,394,717	769	808	5.56 / 5.84
10	133,797,422	728	909	5.44 / 6.79
11	135,086,622	1,279	1,136	9.47 / 8.41
12	133,275,309	1,033	1,304	7.75 / 9.78
13	114,364,328	323	613	2.82 / 5.36
14	107,043,718	701	988	6.55 / 9.23
15	101,991,189	606	1,037	5.94 / 10.17
16	90,338,345	864	1,175	9.56 / 13.01
17	83,257,441	1,185	1,296	14.23 / 15.57
18	80,373,285	268	659	3.33 / 8.2
19	58,617,616	1,468	968	25.04 / 16.51
20	64,444,167	540	602	8.38 / 9.34
21	46,709,983	235	419	5.03 / 8.97
22	50,818,468	487	524	9.58 / 10.31
X	156,040,895	840	654	5.38 / 4.19
Y	57,227,415	63	108	1.1 / 1.89
M	16,569‡	13	24	784.6 / 1448.49

* GRCh38.p11 reference assembly [20]

† GENCODE v27 comprehensive gene annotation set [21]

‡ The revised Cambridge reference sequence of the human mitochondrial DNA (NCBI RefSeq NC_012920.1)

1.4.2 GENES

Genes are the basic units of genetic information. The precise meaning of the term is frequently refined as genetics as a science progresses and our understanding of how genes work improves. According to a modern definition, an individual gene "*is a union of genomic sequences encoding a coherent set of potentially overlapping functional products*" [25]. In practice, the functional products of genes are RNA transcripts. Genes are thus genomic sequences which are transcribed into a related set of functional

RNA molecules. If at least one of the RNA molecules encoded by a gene is a messenger RNA molecule, translated by ribosomes into a protein, the gene is referred to as a protein coding gene, otherwise as a non-coding gene.

There are approximately 20,000 protein coding genes and, according to some estimates, over 23,000 non-coding genes in the human genome (Table 1) [21]. While there is a broad consensus of the identity and amount of human protein coding genes, the number of non-coding genes is much more uncertain. The v27 comprehensive gene annotation set produced by the GENCODE consortium annotates 23,544 non-coding genes, but the NCBI Homo sapiens annotation release 108 contains only 17,835 [25-27]. The wide discrepancy in these estimates reflects the fact that protein coding genes are much more studied and better understood than non-coding genes.

Few mammalian protein coding genes encode a single transcript. Instead, most are transcribed into multiple alternative RNA isoforms, some of which may correspond to different protein isoforms and some which are never translated. As a typical example, consider the human gene *DRAM2*, shown in Figure 2 (panels A-C). The promoter region and transcription start sites (TSS's) of *DRAM2* overlap with those of the adjacent gene *CEPT1* (Panel D). *DRAM2* encodes two well-annotated mRNA's, which have slighty different transcription TSS's. The two mRNA's thus have different first exons, which contain the 5' untranslated regions (UTR's, panel A). The *DRAM2* gene also produces a transcript with a very different open reading frame (ORF), which might be translated into a protein but, according to current annotation, is instead eliminated via nonsense-mediated degradation (Panel B). The exons of *DRAM2* are much more conserved than its introns and appear as spikes on a graph of sequence conservation (Panel F). As most genes, *DRAM2* produces a significant number of processed noncoding transcripts, which do not contain an ORF (Panel C). The biological significance of these, if any, is currently unclear. [21]

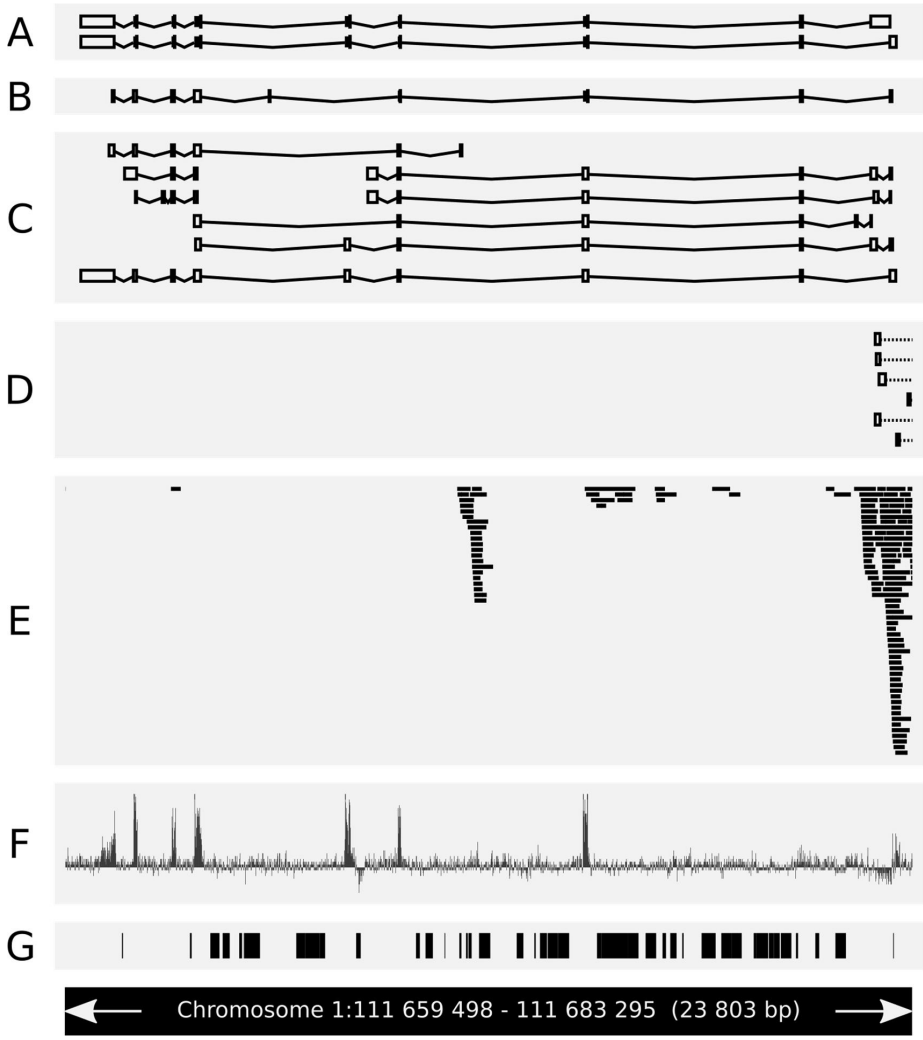

Figure 2 The gene *DRAM2* and its genomic environment. Transcripts are shown as annotated in GENCODE v19, oriented so that the promoter and 5' side of *DRAM2* are on the right [21]. (A) Protein-coding *DRAM2* transcripts. (B) A *DRAM2* transcript degraded via nonsense-mediated decay. (C) Non-coding *DRAM2* transcripts. (D) Transcripts from the adjacent gene *CEPT1*, which partially overlap *DRAM2*. (E) Transcription factor binding sites detected in multiple cell types by the Encyclopedia of DNA Elements (ENCODE) consortium [28]. (F) Sequence conservation based on sequence alignments of 100 vertebrate genomes (G) Repeating DNA elements by RepeatMasker. Tracks E to G extracted from the UCSC Genome Browser on 22.9.2017 [29].

1.4.3 GENE REGULATION

Differences in cell types in multicellular organisms are caused by differential gene expression. Thus, an important portion of the human genome is dedicated to regulating the expression of genes. Some of its mechanisms are

genetic or direcly dependent on specific DNA sequences, but some are epigenetic or unrelated to the DNA sequence itself [30]. In cells, DNA is tightly packed around proteins, most importantly histones, forming chromatin, which is inaccessible to the machinery responsible for transcription. The organization and epigenetic control of chromatin structure, e.g. via DNA methylation and histone modification, play an important role in controlling which genes are actively transcribed. [31]

Two classes of regulatory sequences, promoters and enchancers, are central to driving the expression of eukaryotic genes and the genetic or sequence-specific control gene expression. Promoters ensure that transcription is initiated at the correct position. They contain the core promoter where the basic molecular machinery responsible for synthesising RNA is assembled and which may also bind other transcription factors affecting the level of expression. [12] Enhancers, on the other hand, are distal elements which bind transcription factors, interact with the promoter, and increase gene expression, mostly irrespective to the exact position of the enhancer relative to the target promoter [32].

Many systematic efforts have cataloged and characterized regulatory DNA and epigenetic landscapes in diverse cells. They include strategies such as screens of transcription factor binding sites, histone modifications, and DNA methylation, as well as screens of chromatin accessibility (Panel E, Figure 2). [33,34] Although advances have been made in identifying e.g. enhancer elements, comprehensive understanding of gene regulation still requires significant progress [35]. The function of distal regulatory elements in the human genome is particularly poorly understood.

1.5 HUMAN GENETIC VARIATION

SNPs are numerically by far the largest class of genetic variants in humans with millions of SNPs identified to date (on 10.2.2018 the NCBI dbSNP database includes 325.7 million reference SNP ID numbers) [36]. Humans carry a SNP allele at an average of 1,500,000 sites per individual. [6] Due to their high number and relatively easy accessibility to different laboratory

methods, SNPs have gained a prominent position in human genetic research, where the term SNP is commonly used as a synonym of small or short genetic variants in general. The other variant classes tend to receive less attention, although CNVs, for example, encompass a larger proportion of the genome than SNPs [13].

The overwhelming majority of variant alleles constituting human genetic variation, the genetic differences between humans, are not new mutations. A typical human genome differs from the human reference genome at 4.1 million to 5.0 million sites, but carries, on average, less than 80 *de novo* mutations [6,37]. Human genetic variation is therefore, by and large, related to alleles inherited from previous generations. Thus, the characteristics of human genetic variation are mostly the result of human population history and forces acting on historical timescales, such as migrations and population growth. For example, in a sample of African Americans, the average age of variant alleles in protein coding genes was estimated to be 47,600 years [38].

1.5.1 GLOBAL POPULATION HISTORY AND GENETIC VARIATION

Both paleontological and genetic evidence place the lower limit for the age of *Homo sapiens* as a species to roughly 300,000 years [39-41]. According to current understanding, anatomically modern humans spread out of Africa to inhabit the rest of the world some 100,000 to 50,000 years ago [42,43]. The number of individuals whose descendants first populated regions outside Africa was quite small, corresponding to an effective population size of approximately 1,000 to 5,000 [44,45]. With global population size increasing, new territories were populated during the following millennia by select groups of people, giving rise to local subpopulations [46,47].

When humans were living only in Africa, there were on the order of 100,000 to 500,000 people in the world [45,48]. Following the colonization of the world outside Africa, global population size increased notably, and the growth rate accelerated roughly five-fold some 10,000 years ago, coinciding with the invention and adoption of agriculture [49]. Growth then stayed roughly constant until the population explosion brought by the industrial

revolution in the 19th century, when the world's population grew from 1 billion to more than 7 billion in just two centuries [50,51].

The way the globe was populated left important marks on the genetic variation of present-day humans. The entire process can be roughly modeled as a series of founder effects originating in Africa, with some migration between the resulting populations [46]. A founder effect, in which a relatively small number of individuals or founders give rise to a new population, is characterized by a loss of genetic diversity and the outsize frequency of the alleles carried by the founders in the subsequent generations. Furthermore, as rare alleles are less likely to be carried by the founders and more likely to be lost due to genetic drift in small populations, the rarest alleles are most likely lost during and following a founding event, when the population size is not large. [52] Non-African populations have thus inherited a subset of the genetic variation found in Africa, so that each successive founding event came with a corresponding loss of genetic diversity with rarer alleles lost more often than more common ones. [44,46]

Contrary to the disappearance of rare alleles in founding events, population growth leads to an accumulation of rare alleles. Importantly, much of the growth of human populations has occurred recently, after the colonization of the world outside Africa. This recent population growth has thus increased the number of rare alleles in human populations but, as the growth occurred after the populations had diverged, many of these alleles are not shared between the populations. [6,53] Together, these two forces, the ancient serial founding effects out of Africa and the relatively recent explosive population growth, have created the two defining features of global genetic variation of humans: First, alleles which are common in any given non-African population are very likely old and common in others as well (Figure 3) [6,54]. Second, there is a multitude of younger rare alleles, many of which are population-specific (Figure 3; Figure 4) [53].

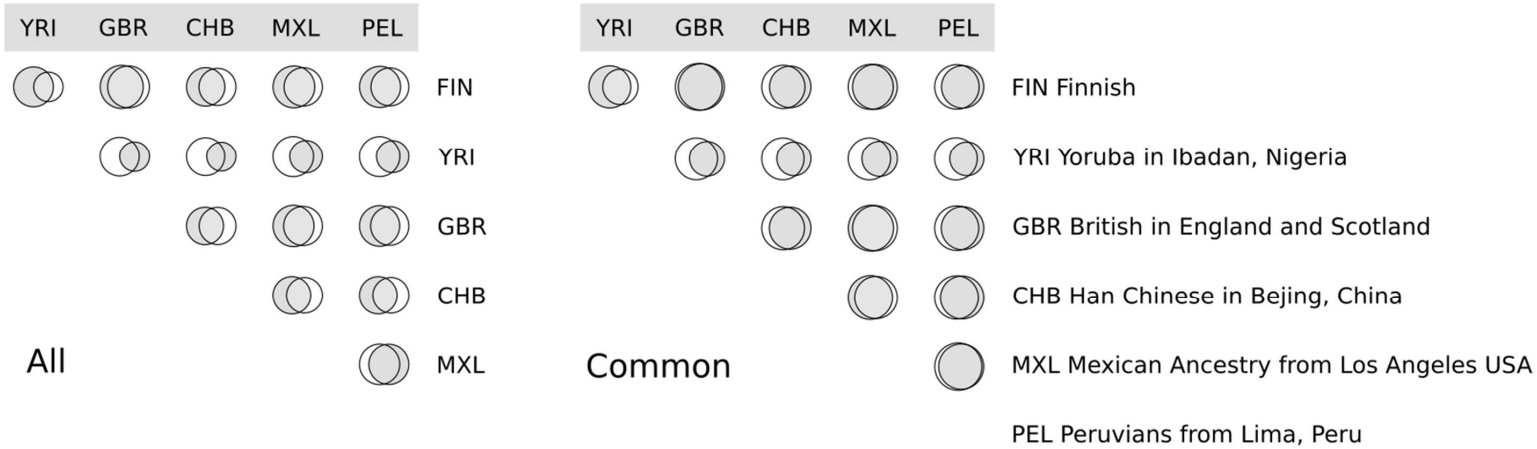

Figure 3. Sharing of variant alleles in whole-genome sequencing data from the 1000 Genomes Project Phase 3 variant calls on chromosome 1 [6]. For each variant position, the global minor allele is set as the depicted alternate allele. For each pair of populations, the area of the circles is proportional to the number of different alternate alleles found in the population, and the area of the intersection is relative to the number of alternate alleles found in both populations. Circles on the left show all variant alleles, while those on the right include only alleles with MAF > 0.05 in each population. Note that the sample sizes are on the order of approximately 100 participants per population. With larger sample sizes, the number of rare alleles would have been even greater, while the number of common alleles would not have substantially changed.

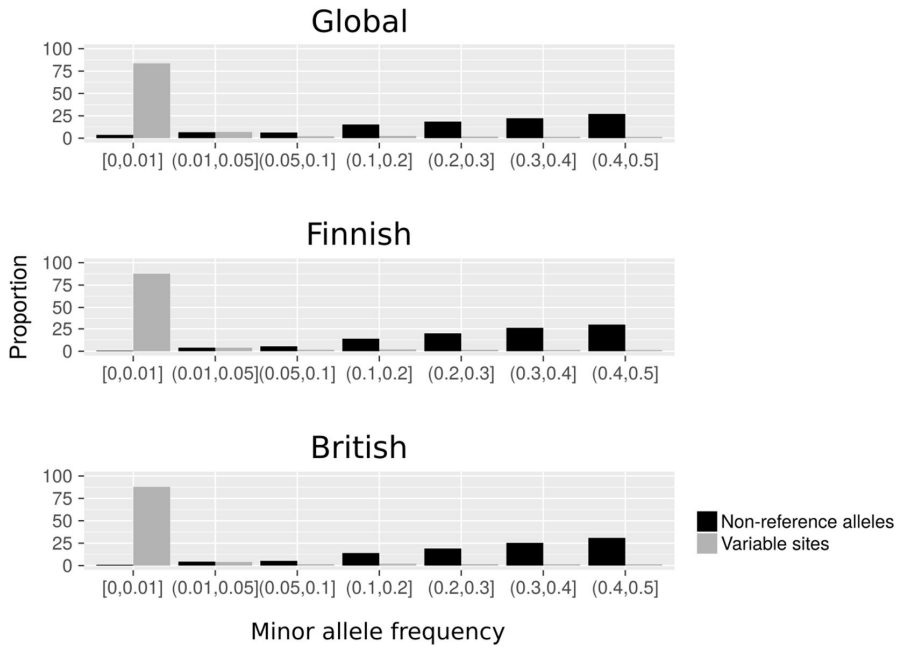

Figure 4 Allele frequency distribution and genetic variation. The gray bars show the proportion of all variable sites according to their minor allele frequencies. The black bars show the same proportion weighted by the number of non-reference alleles carried across the study participants. Data from whole-genome sequencing data from the 1000 Genomes Project Phase 3 variant calls on chromosome 1 [6].

1.5.2 POPULATION HISTORY OF FINNS AND GENETIC VARIATION IN FINLAND

The colonization of Finland was one end of the serial founder events originating out of Africa. Situated at the northern edge of the habitable world, Finland was covered by ice until the end of the last glacial period some 10,000 years ago. As the ice sheet retreated, nomadic hunter-gatherers explored the lands and were the only inhabitants for millennia. [55] It is thought that most of what would later be called the ancestors of modern Finns arrived much later, some 4,000 years ago, with another significant migration into Finland approximately 2,000 years ago [5]. The number of these founders has been estimated to be quite small. For most of the time,

since the arrival of the first migrants to Finland, they inhabited only the southernmost coastal area of the country. [5,56]

The population size of Finland stayed small for centuries after the migration waves 4,000 and 2,000 years ago. In the 12th century, there were only an estimated 50,000 Finns [5]. The small number of founders, the resulting small population size of Finland, and the very slow growth in population size lasting for centuries created a genetic bottleneck. This lead to a substantial loss of genetic variation in Finns relative to its neighboring countries [56]. Many rare alleles were entirely lost, leading to a general deficit of rare alleles in Finns relative to most other Europeans [57]. Conversely, many of the rare alleles which survived the bottleneck increased in frequency becoming much more common. As a result, there are approximately at least 30,000 alleles which are relatively common in Finns (MAF>5%), but rare (MAF<0.5%) in other populations, including non-Finnish Europeans, although this estimate is limited by the small number of Finns surveyed [6].

In the 16th century the population of Finland started to grow, and regions outside the coastal areas started to be more extensively populated. Isolation by geographic distance was, however, strong and with limited internal migration. The treaty of Nöteborg (Pähkinäsaaren rauha) in 1323 established the first official border in Finland, then a part of the Swedish kingdom. This appears to have created a permanent impediment of internal migration and left an important imprint in the Finnish population with marked differences between Finns living in the south and west as compared to those living in eastern or northern Finland. [58-60]

1.6 GENETIC VARIATION IN TRAITS

1.6.1 HERITABILITY

In attempting to understand the causes of variability in a trait, such as susceptibility to a disease, it is helpful to be able to divide the variability into variability caused by genetic and non-genetic causes. The concept of heritability formalizes this division and can be used to quantify the

proportion of total variability that can be accounted for by genetic factors. [61,62] The total phenotypic variance in a trait can be divided into variance attributable to genetic and non-genetic or environmental factors as

$$V_P = V_G + V_E$$

where V_P, V_G, and V_E denote the total, genetic, and environmental variances, respectively. The genetic variance can be further divided into additive, dominance, and interaction effects V_A, V_D, and V_I:

$$V_G = V_A + V_D + V_I$$

Given these definitions, the proportion of phenotypic variability caused by genetic effects or the broad sense heritability of the trait H^2 is given by

$$H^2 = \frac{V_G}{V_P}$$

while h^2, defined as

$$h^2 = \frac{V_A}{V_P}$$

and known as the narrow sense heritability, is the proportion attributable to additive genetic effects. [63]

Heritabilities have been estimated for hundreds of human traits using various different methodologies. [63,64] According to a meta-analysis of 2,748 publications estimating heritabilites using the classical twin study design, for almost all traits the narrow sense heritability is nearly equal to the

broad sense heritability with a mean value of h^2=0.488 across the 17,804 traits included in the analysis [64]. Thus, as a rule of thumb, roughly half of the variance of any given human trait may be expected to be caused by genetic variation.

1.6.2 PENETRANCE AND EFFECT SIZE

The phenotypic consequences of genetic variation range from non-existent to lethal physiological defects. The vast majority of variant alleles in humans have no detectable phenotypic effects. Most of those that affect the phenotype are neutral in terms of natural selection [65]. Furthermore, many of those alleles that may change the phenotype often do so only under certain circumstances [66].

The probability that an individual carrying a given genotype will express a specific qualitative phenotypic effect is called the genotype's penetrance [67]. Penetrance is not a function of only the biological properties of a given allele. Instead, genetic and environmental backgrounds are important factors defining the penetrance of the genotypes of many alleles. Importantly, the penetrances are at least in theory modifiable: A successful treatment of a genetic disease reduces the disease allele's penetrance. Likewise, a population-based intervention targeting a disease's environmental risk factors may positively affect the penetrances of the alleles contributing to the disease's prevalence.

For quantitative traits, the effect of an allele needs to be expressed relative to an arbitrarily selected reference genotype. The effect can then be stated in terms of the mean phenotypic difference between the genotypes, such as the mean difference in blood pressure between heterozygotes and reference homozygotes. For quantitative traits, effect sizes are often expressed relative to the sample standard deviation, which enables comparisons across different samples and phenotypes measured in different units such as body mass and height.

Analogous to quantitative traits, the effects of alleles contributing to qualitative traits can also be expressed relative to a reference genotype. For

disease traits, two related measures, the relative risk (RR) and the odds ratio (OR) are often used. Given a disease D, a genotype *i* and a reference genotype *r*, they are defined as follows [67]:

$$RR_i = \frac{P(D|i)}{P(D|r)}$$

$$OR_i = \frac{P(D|i) \,/\, \big(1 - P(D|i)\big)}{P(D|r) \,/\, \big(1 - P(D|r)\big)}$$

The relative risk RR is thus easily understood as the ratio of the penetrances of a genotype and a reference genotype. While there is no similarly intuitive interpretation for the odds ratio, it has useful statistical qualities making it often used in practice. Notably, the odds ratio is directly related to the regression coefficient of an independent variable in a logistic regression model [68].

1.6.3 MONOGENIC TRAITS AND DISEASES

Some diseases or traits are caused by a single disease allele. These monogenic diseases are often rare and severe. [12,69,70] Alleles causing monogenic diseases have typically high penetrance and, if the diseases are severe, the frequency of the disease alleles is low as a result of purifying selection against the alleles [69-71].

The inheritance of many monogenic diseases displays dominance, i.e. the phenotype or the penetrance of the heterozygous genotype is very close to one of the homozygotes, rather than being at an intermediate point between the two [12,70]. In the case of recessive monogenic diseases, the heterozygotes carrying only one copy of the disease allele do not have the disease. Conversely, in dominant recessive diseases, the heterozygotes are affected with the disease. These modes of inheritance and their unique segregation patterns in pedigrees were described by Mendel based on his

experiments with plants. Because of this, monogenic diseases are often also called Mendelian diseases.

Although an important number of monogenic diseases remain poorly understood, Mendelian disease genetics has been very successful in identifying specific disease alleles and genes [69]. A notable example of this success is the identification of disease alleles responsible for a group of monogenic diseases, which are much more common in Finland than elsewhere in the world, constituting what was named as the Finnish disease heritage by the pioneering Finnish geneticist Reijo Norio [5].

1.6.4 COMPLEX TRAITS AND DISEASES

Many diseases and phenotypes are determined by the alleles of many genes together with environmental factors. These complex or multifactorial traits include most familiar phenotypes and diseases, such as height, body mass, blood pressure, and coronary artery disease. Complex diseases, as opposed to monogenic diseases, are typically common in the population, have a late disease onset with a strong component of environmental or life style factors affecting the disease risk. [12]

At the turn of the 21st century, some scientists engaged in a debate regarding the genetic structure or architecture of common diseases. In theory, many different genetic models are compatible with the observed characteristics of late-onset common diseases, such as coronary artery disease. The central point of the debate concerned the frequency of the disease risk alleles. They could be quite common in the population, in which case they necessarily would have, at most, modest penetrances to be compatible with the observed prevalences of said diseases. Conversely, there could be a multitude of rare disease alleles, even if the diseases themselves were relatively common. If the disease alleles were rare, the observed disease prevalences would be compatible with many possible penetrance distributions of the rare alleles. [72,73]

A specific study design, the genome-wide association study (GWAS), was suggested based on the idea of the existence of common disease risk alleles. [73,74] This study design has since been used to investigate a wide variety of

complex phenotypes, including common complex diseases. As a partial confirmation of the "common disease, common variant" hypothesis, a general theme for the various phenotypes studied is that complex traits have a high number of common contributing alleles. However, the effect sizes of these alleles are much smaller than anticipated, with the heterozygous odds ratios typically ranging from 1.1 to 1.4, and their number has also been greater than expected by many. [75,76]

1.7 GENETIC MAPPING OF TRAITS

Identifying genes and genetic variants affecting the occurrence of traits and diseases has both intrinsic value as a way to advance our biological understanding and instrumental value as a means to develop e.g. diagnostics, therapeutics, or other interventions for disease. Genetic mapping,or determining the locations of genes or alleles correlated with traits, can be roughly divided into studies where the genetic makeup of an organism or cells is artificially manipulated and studies investigating the phenotypic correlations of naturally occurring genetic variation.

Studies of humans are, for obvious reasons, limited to studying either naturally occurring genetic variation or cell models. Many traits and diseases, however, are not easily reproduced using cellular models. Furthermore, developing a useful cellular model typically requires some understanding of the relevant disease processes, which is often lacking. Studying the effects of naturally occurring genetic variation is thus the primary means of genetic mapping in humans.

1.7.1 THE P-VALUE AND STATISTICAL POWER

The success of a genetic association study depends on its ability to declare true associations as statistically significant. This concept, known as statistical power, is defined in relation to the null hypothesis H_o. In the context of genetics, H_o includes the assumption of no dependence or association between the genotypes and the trait under study. H_o is used to justify a statistical model of the data, which relates the observations to a random

variable with a known distribution. The value of the random variable corresponding to the observations, known as the test statistic, is compared to its distribution to calculate the probability of the test statistic being at least as extreme as its value calculated from the data. This probability is known as the test's P-value. If the P-value is smaller than some value α, the null hypothesis is rejected and the test is declared statistically significant. Setting the value of α therefore defines how often it is acceptable to declare tests statistically significant when the null hypothesis is true. Scientific literature commonly accepts the value $\alpha=0.05$.

Depending on whether H_0 is true or false, two types of correct statistical decisions and two types of incorrect decisions can be made. Type 1 error, or false positive results, rejects H_0 when it is true. Type 2 error, or false negative results, do not reject H_0 when it is not true. True positive results reject H_0 when it is not true, and true negative results do not reject H_0 when it is not true (Table 2).

The power of a statistical test is the probability of getting a P-value smaller than α when the null hypothesis is not true. The power of a study is often estimated to assess whether a study design is able to produce meaningful results. Various parameters need to be determined (e.g. effect size and LD between the causal variant and the marker), after which desired statistical models can be used to produce power estimates.

Statistical hypothesis testing using P-value cutoffs has been criticized, as it is somewhat arbitrary and often misused or misunderstood. Some have advocated abandoning it entirely, while others suggest changing the accepted cutoffs to more stringent values. [77,78] However, at least for the time being, most biomedical journals and reviewers of scientific papers in practice require the use of P-values.

Table 2. Statistical testing.

Decision	Reality	
	No association	Association
Reject H_0	False positive (Type 1 error)	True positive
Do not reject H_0	True negative	False negative (Type 2 error)

1.7.2 GENETIC LINKAGE AND ASSOCIATION

Establishing a connection between an allele and a trait eventually boils down to studying their co-occurrence. Two general types of tests have been used in human genetics: tests of linkage and tests of association. Genetic linkage studies investigate the co-occurrence of transmission of alleles and the phenotype from parents to their offspring in pedigrees. As the studies constituting this thesis did not employ tests of genetic linkage, this class of tests is not discussed further here. Genetic association studies, on the other hand, compare the frequencies of alleles between classes of individuals (e.g. affected vs. unaffected) or trait values between genotype classes (e.g. the height of homozygotes vs. heterozygotes). [67] To explicitly denote that genetic variants assayed in typical genetic mapping studies are often not expected to be the actual causal variants, they are called marker loci or genetic markers.

1.7.3 GENOME-WIDE ASSOCIATION STUDIES

The first genetic association studies were strongly hampered by the poor technological capacity to determine the genotypes of genetic variants. Until roughly 2005, it was only feasible to asses the genotypes at a few variant positions. Scientists thus tried to guess which genes might have something to do with their trait of interest and selected the few genetic variants in some way near these genes. The association studies of this era, collectively called candidate gene studies, are notorious for their lack of robustly replicable results, especially for complex phenotypes [79-82].

The idea that common genetic variation might be responsible for much of the variation in many interesting traits had important implications for

genetic association studies. Non-African populations have a limited number of common alleles, which are mostly shared across these populations. Furthermore, as the out-of-Africa expansion happened recently (in historical terms), there has been relatively little time to for recombination to eliminate the linkage between the alleles. As a result, linkage disequilibrium (LD) is extensive throughout the human genome. Together, these artefacts of human population history made it economically possible to use DNA microarray technology to genotype a sufficiently large number of variants to reach genome-wide coverage of common genetic variation, eliminating the need to select candidate genes or variants for association studies [83]. Relying on LD, the result of distant relatedness, in effect exploits the giant pedigree connecting all humans. This is similar to what is made in genetic linkage studies, but without actually modeling the unknown shape of the pedigree and the transmission of alleles within it (Figure 5) [67,84,85].

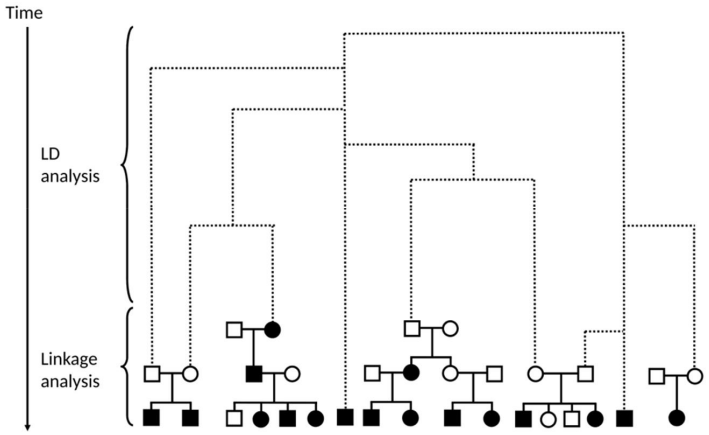

Figure 5. Association analysis based on LD in the unknown pedigree compared to linkage analysis in known pedigrees. Reproduced with permission from Tero Hiekkalinna [85].

A catalog of common human genetic variation and an estimate of its correlation structure is necessary for the design of DNA genotyping microarrays that can be used in GWA studies. The International HapMap consortium produced the first such catalog, partly utilizing other data

produced prior to the project [54]. Later, the 1000 Genomes Project and the Haplotype Reference Consortium have collected larger and more comprehensive catalogs [6,86]. Current efforts expand the coverage of the catalogs to variants with lower frequencies and new human populations. The GWA genotyping microarrays almost exclusively assay SNPs, as these are both numerous and relatively easy to genotype.

In genome-wide association (GWA) studies, a given statistical test of association between a genetic variant and a trait is independently repeated for all of the genotyped variants. Linear and logistic regression models are most often used, as they allow for including covariates in the test. The genetic variant is typically modeled as a 0, 1, 2 coded variable, which corresponds to an additive genetic model in simple linear regression and to a multiplicative genetic model in logistic regression.

Association studies in general and GWA studies in particular are mostly made using population-based samples of study participants who are assumed not to be closely related to each other. [75] Although genetic association tests can be made in any type of study samples, including family-based samples and pedigrees, it is much easier to recruit a number of unrelated individuals than to track down similar numbers of people from a single or a few families. Thus, the typical setting for GWA studies is a population-based cohort. For disease traits, these cohorts are often collected as case-control cohorts of affected and unaffected study participants.

The high number of genetic markers used in GWA studies needs to be taken into account when defining the statistical significance of the results. Different types of adjustments to the commonly accepted limit of statistical significance (P=0.05) have been proposed and used. In practice, two are most often employed: either multiplying the P-values by the number of tests made (known as the Bonferroni correction) and using the commonly accepted P<0.05 statistical limit of significance for the Bonferroni-corrected P-values, or the fixed $P<5\times10^{-8}$ limit of genome-wide significance. The latter was first put forward as a crude estimate in a seminal article by Risch and Merikangas [73]. It was later derived by approximating the theoretical number of independent tests equivalent to testing all common human

genetic variants, taking into account the non-independence of the real genetic variants due to LD [87].

A practical consequence of the large number of tests made in GWA studies is that it nearly guarantees that any systematic biases have plenty of opportunities to appear in the data. For example, if there are slight differences in how the DNA samples from affected and unaffected study participants are handled, it is likely that they will have an effect on *at least some* of the thousands of genetic markers studied, possibly creating artefactual signals of association. Thus, a reasonable study design, including how the DNA samples are processed and genotyped, and a robust quality control procedure are required in order to produce valid results.

1.7.4 GENOTYPE IMPUTATION

Modern genotyping arrays are able to query hundreds of thousands of variants. The resulting data can be further improved by using genotype imputation to infer the genotypes at untyped genomic positions and positions where genotype was not successfully determined using the microarray. One of the benefits of genotype imputation is that it harmonizes and standardizes data produced on different genotyping arrays, which often genotype different sets of markers. This enables the combination of multiple datasets using meta-analysis, resulting in larger effective study samples. [88]

Genotype imputation relies on an external reference dataset consisting of a set of known (or phased) haplotypes which, together with the genotyped study data, are fed to an imputation algorithm. Imputation algorithms then predict the genotypes at those positions which were not present in the study data based on the reference data. Publicly available reference haplotype datasets include those produced by the HapMap, 1000 Genomes, and the Haplotype Reference Consortium projects [6,54,86]. A number of efficient imputation algorithms have also been developed, making genotype imputation a routine part of most GWA studies. [88-90]

1.7.5 LIMITATIONS

There is an important limitation regarding the utility of results from genetic mapping studies, which is exacerbated in studies of common traits or common genetic variation: The extensive LD present in human populations makes it very hard, if not impossible, to attribute associations to single causal alleles, as alleles in any given genetic region are intercorrelated. In the case of rare disease alleles with very high penetrances, often located in protein-coding regions, this is easier, as the causal allele or genotype is almost never seen in healthy individuals. In the case of common low-penetrance alleles, however, most of those who carry the disease allele or genotype do not have the disease. To make the situation even more difficult, most associations identified using GWAS are related to regulatory variation, not variation in protein structure. Currently, our understanding of regulatory DNA in the human genome is far from sufficient to mechanistically understand regulatory variation. As a result, the results of GWA studies mostly only indicate broad genomic regions as associated with a trait, but do not unambiguously indicate specific genes or alleles. Scienctists typically try to conjure *ad hoc* solutions to this problem, with varying degrees of success. Solving this problem is among the most important current challenges facing human geneticists. [91]

1.8 HEART FAILURE

1.8.1 DEFINITIONS

Heart failure (HF) is a clinical syndrome that may be caused by a number of different factors, most commonly ischemic heart disease, hypertension, and valvular heart disease [2,4]. The different causes lead to a final common pathway where similar mechanisms are involved, partially independent from the aetiology [92]. The current guidelines of the European Society of Cardiology (ESC) define heart failure as "*a clinical syndrome characterized by typical symptoms (e.g. breathlessness, ankle swelling and fatigue) that may be accompanied by signs (e.g. elevated jugular venous pressure, pulmonary crackles and peripheral oedema) caused by a structural and/or*

functional cardiac abnormality, resulting in a reduced cardiac output and/or elevated intracardiac pressures at rest or during stress" [4]. Essentially, heart failure is a significant deficit in the heart's pumping capacity, causing inadequate circulation.

Being a complex clinical syndrome, heart failure can be divided into a plethora of categories. The different classifications are based on its suspected aetiology, different physiological parameters, clinical features, temporal stage and the structural properties of the heart. [4] The classifications are overlapping and many classifications are often simultaneously used to describe different types of heart failure [3,4,93].

A central classification of heart failure is based on left ventricular ejection fraction (LVEF), the percentage of blood pumped out from the left ventricle with each systole. In heart failure with preserved LVEF (HFpEF), the percentage of blood pumped out is considered normal (LVEF ≥ 50%), whereas in HF with reduced LVEF (HFrEF) it is smaller than usual (LVEF ≤ 40%). Those cases falling in between the two cut-offs may be classified as HF with mid-range EF (HFmrEF). [4] HF with preserved and reduced LVEF are also often called somewhat ambiguously diastolic and systolic HF, respectively, but diastolic and systolic abnormalities in HF are not mutually exclusive [4,93].

1.8.2 PREVALENCE

In the vast majority of cases, heart failure is a common late-onset syndrome. The prevalence of heart failure, or the number of affected individuals, has been consistently estimated to be roughly 1-2% in developed western countries. [93] The prevalence increases sharply with age, reaching approximately 7-8% in those older than 75 years [93-95]. Data from the developing world are scarce, but heart failure is definitely not a problem of the affluent western countries alone [3,96,97].

1.8.3 SYMPTOMS AND DIAGNOSIS

Typical symptoms of heart failure include dyspnea or breathlessness, even when lying flat (orthopnea), fatigue and ankle swelling. However, the signs and symptoms may be quite non-specific, ranging from nocturnal cough to depression. Diagnosis of HF is therefore often based on ruling out suspected heart failure and requires the presence of symptoms, clinical findings, and objective evidence of cardiac alterations underlying HF. The ESC suggests a diagnostic algorithm which uses the circulating concentrations of B-type natriuretic peptide (BNP) or N-terminal pro-BNP (NT-proBNP) as a biomarker to rule out suspected HF (Figure 6). [4]

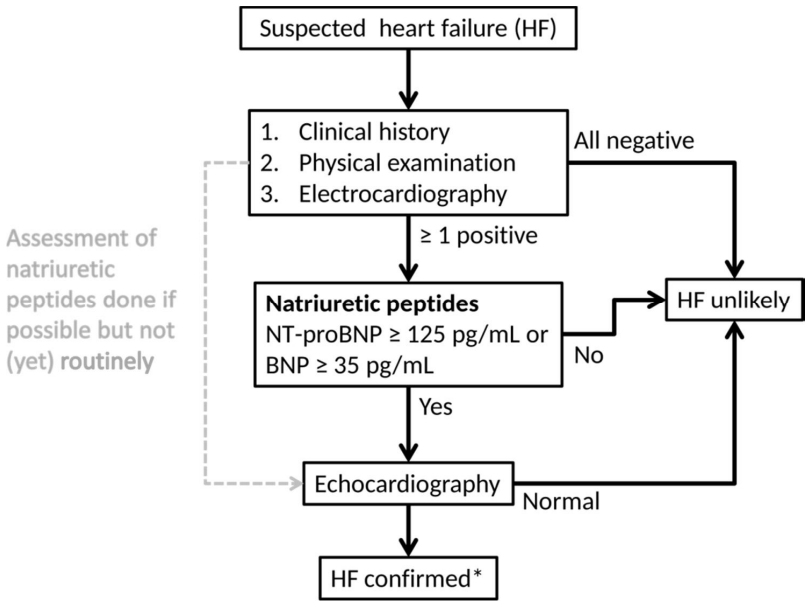

Figure 6. A diagnostic algorithm for the diagnosis of suspected heart failure as suggested by the ESC. Modified from [4]. *The ESC guidelines explicitly allow for uncertainty even at the end of the diagnostic algorithm, indicating that the diagnosis is or should be confirmed "based on all available data".

1.8.4 NATRIURETIC PEPTIDES IN HEART FAILURE

The atrial natriuretic peptide (ANP) and B-type natriuretic peptide (BNP) and small hormones are produced by cardiomyocytes when they are physically stretched. They were discovered relatively recently, in 1985 and 1988 [98,99]. Atrial cardiomyocytes mostly secrete ANP, while ventricular

cardiomyocytes produce predominantly BNP [100]. Both are created as prohormones which are cleaved to release a biologically inactive N-terminal fragment and the C-terminal hormone. For ANP, the cleavage is done by the cardiomyocytes, which secrete the N-terminal fragment and the active ANP into the circulation as separate entities. ProBNP is cleaved after entering the circulation. The main biological functions of ANP and BNP are thought to be the reduction of cardiac load by increasing natriuresis, inducing vasorelaxation, and other physiological effects. [101]

ANP, BNP, and their N-terminal fragments are unique biomarkers of heart failure [4,101-103]. During excessive myocardial stress, their production increases dramatically, even by multiple orders of magnitude. Thus, their normal circulating concentrations are not compatible with existing heart failure and they can be used as rule-out criteria. However, high concentrations of circulating natriuretic peptides may be caused by factors other than heart failure, most often old age, kidney failure, and atrial fibrillation, so their elevated concentration is not recommended as a means to establishing a diagnosis with heart failure. [4,104]

Prior to this thesis, no estimates of the heritabilities of circulating ANP or BNP concentrations have been published. Candidate gene studies of ANP and BNP and GWA studies of BNP or NT-proBNP have reported associations at five genomic loci, showing that their circulating concentrations are indeed affected by genetic variation [105-110]. The studies have, however, been limited in their genomic and phenotypic coverage.

1.8.5 AETIOLOGY

Heart failure may be caused by a number of very different conditions. The ESC divides them into those affecting the heart muscle itself, abnormalities in the pressure and flow in and through the heart ("abnormal loading conditions"), and arrythmias disturbing the electrical functioning of the heart [4]. The conditions can also coexist and one may trigger the onset of another. [111,112] Quite commonly (even up to 30-50% of the cases by some estimates), a single underlying cause of heart failure cannot be unambiguously identified [112,113].

In the majority of cases, heart failure has a multifactorial aetiology. The two main causes, coronary heart disease and hypertension, are extremely prominent and interrelated. [2,114] Together, they account for more than 50% of heart failure cases in all major regions in a recent survey of Africa, Asia, the Middle East, and South America [97]. In the developed western world, these two factors are even more central. Coronary artery disease was the underlying cause in 62% of the participants in a survey of 24 multicenter HF treatment trials [115]. An estimate from a western population-based cohort suggests that total population-wide elimination of hypertension alone would have reduced the prevalence of heart failure by 39% in men and by 59% in women [2]. Valvular heart disease, where the function of the heart's valves is impaired, is also a notable cause of heart failure, although its relative contribution to the prevalence of HF is smaller than that of coronary artery disease or hypertension [116]. A set of four major modifiable risk factors, including blood cholesterol, smoking, diabetes mellitus, together with hypertension, has an outsize impact on cardiovascular disease in general, including multifactorial heart failure [117].

1.8.5.1 Hypertension

Arterial hypertension or high arterial blood pressure is commonly defined as either systolic arterial blood pressure (BP) greater than 140 mmHg or diastolic arterial BP greater than 90 mmHg [118]. Its heritability has been estimated to be in the range of 30-70%, depending on the study [118,119]. Essential or primary hypertension, which has no specific identifiable cause, accounts for up to 95% of the cases [120]. However, strong epidemiological evidence points to excess intake of dietary sodium, typically associated with urban or western diets, as the main culprit in hypertension [118,121]. Hypertension is practically absent in remote rural populations where the intake of dietary sodium is low [122]. It is therefore highly probable that HT in general, as well as genetic variation in HT susceptibility, are ultimately related to sodium intake and its physiological effects.

In the absence of a myocardial infarction, hypertension can lead to the onset of heart failure via complex changes in the structure of the heart. These

changes are related to the need for increased pumping force required to overcome the high blood pressure. They include ventricular hypertrophy, where cardiomyocytes increase in size (analogous to the hypertrophy of skeletal muscle in excercise) and dilation of the ventricles. The events and mechanisms of how these changes eventually lead to heart failure have been the focus of extensive studies but remain poorly understood. [123]

Arterial blood pressure, being both of high medical importance and relatively easy to measure, is one of the complex traits most studied by human geneticists. Linkage studies of rare Mendelian forms of hypertension were succesful in identifying rare genetic variants, mostly in kidney-related genes. However, initial studies assessing typical population variation in BP with genome-wide linkage studies and candidate gene association studies were largely a failure until quite recently. [124] Finally, in landmark studies published in 2009 and 2011, large GWAS meta-analyses robustly identified several genomic regions associating with blood pressure in the general population [125-127]. Interestingly, and in contrast to the variants identified in Mendelian HT, these variants appear to be largely unrelated to kidney function [127].

1.8.5.2 Coronary artery disease and myocardial infarction

Coronary artery disease (CAD), leading to coronary heart disease (CHD), is a major cause of heart failure. It is a slow, chronic, inflammatory process comprising the formation of lipid-rich plaques inside the walls of coronary arteries. The plaques trigger immune cells, particularly macrophages, to infiltrate the coronary wall towards the plaques. A coronary plaque may rupture or erode, exposing thrombotic material from inside the plaque to circulation and causing the formation of a thrombus, also known as a blood clot. The thrombus in turn may disrupt the flow of blood, leading to the onset of acute coronary syndromes and ischaemic damage to the part of the heart which depends on the blood flowing through that particular artery. [128]

Acute coronary syndromes (ACS), comprising acute myocardian infarction and unstable angina pectoris, are caused by a decrease in coronary blood flow [129]. The ischemia is often severe enough to lead to the death of

cardiomyocytes or myocardial infarction (MI). More rarely, the ischaemia causes chest pain or angina pectoris but no biochemically detectable death of cardiaomyocytes, in which case the ACS is labeled as unstable angina pectoris (UA). MI is further divided into ST-elevation MI (STEMI) and non-ST-elevation MI (NSTEMI) based on specific electrocardiographic (ECG) patterns (Figure 7) [129,130]. MI caused by the disruption of a coronary plaque is also collectively denominated MI type 1, in contrast to MI type 2, where the ischemia is not caused by CAD but other factors, such as a coronary embolism [131].

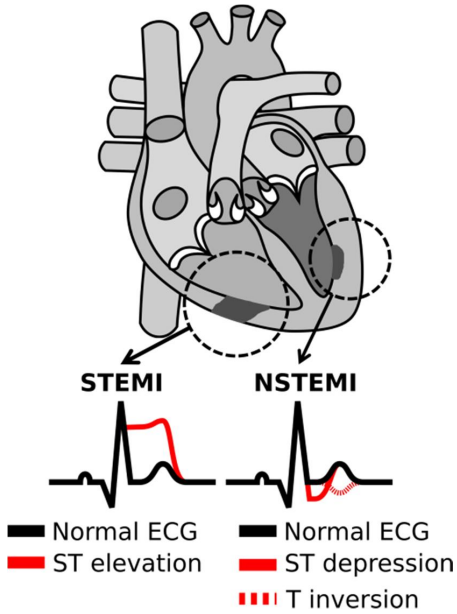

Figure 7. ST-elevation and non-ST elevation myocardial infarction. Transmural infarction, spanning the entire thickness of the ventricular wall, is reflected as an elevation of the ST-segment in an electrocardiogram. A subendocardial infarction, with myocardial injury restricted to myocardium more distant from the blocked artery, may result in no elevation of the ST segment, its depression, or T-wave inversion. Figure modified from https://commons.wikimedia.org/wiki/File:Diagram_of_the_human_heart.svg under the license available at https://creativecommons.org/licenses/by-sa/3.0/legalcode.

CAD may lead to heart failure via myocardial injury related to ischemia in ACS. As the proliferative capacity of cardiomyocytes in adults is essentially absent, myocardium destroyed by an infarction cannot regenerate [132]. The healing of damaged myocardium is therefore limited to the formation of a fibrotic scar and a large infarction can directly lead to heart failure simply

due to the quantity of cardiomyocytes lost [133]. Heart failure may also gradually develop following an infarction. As the surviving myocardium tries to compensate for the loss of strength, it is excessively stressed, sometimes leading to the onset of heart failure, similar to the formation of hypertensive heart failure. [134]

Although the pathological mechanisms involved in the formation of coronary plaques are complex, the development of CAD is simpler from an epidemiological point of view. The formation of coronary plaques is highly dependent on the circulating concentration of blood lipids, especially cholesterol contained within low-density lipoprotein (LDL) particles. [117,128,135] Individuals homozygous for a loss-of-function mutation in the low-density lipoprotein receptor gene *LDLR*, causing familial hypercholesterolemia, have a highly elevated risk of CAD and its complications [136]. Conversely, LOF mutations in *PCSK9*, leading to reduced cellular degradation of LDL receptors and corresponding low circulating LDL cholesterol concentration, protect their carriers from CAD [137]. Importantly, diet has been robustly shown to be an important modifier of blood cholesterol levels [138,139]. Thus, similar to the effect of dietary sodium intake in hypertension, excess dietary intake of cholesterol is responsible for a large proportion of CAD cases.

The heritability of CAD, estimated to be roughly 40-60%, is highly similar to that of hypertension and many other complex phenotypes [140]. Some cases are essentially monogenic, related to familial hypercholesterolemia, but the overwhelming majority of cases is related to multifactorial or complex CAD [140]. Several GWA studies of CAD have been published. Notably, an association of a common variant on chromosome 9 with CAD was one of the first major results of GWAS [141]. Since then, larger meta-analyses have led to the identification of over 50 additional independent loci [140,142-145]. Emblematic to the difficulty in attributing the results of genetic association studies to specific causal alleles, the functions and mechanisms of most of these associations are unclear. Some of the associated loci can be linked to cholesterol metabolism, but their majority does not associate with any of the known common CAD risk factors. [140,146,147]

1.8.5.3 Takotsubo cardiomyopathy

In the 1990's, Japanese scientists described a peculiar type of transient heart motion abnormality, disturbing the pump function of the heart. The symptoms were highly similar to acute coronary syndromes but appeared in the absence of significant coronary artery occlusion. [148,149] Various names for the disease, such as stress cardiomyopathy, have been suggested after its initial description. Takotsubo cardiomyopathy (TTC), the name given by the Japanese scientists, has however been established in the literature. [149,150] TTC often presents with reduced LVEF and has been labeled as an acute heart failure syndrome by some authors [151].

Since TTC has become more widely known, it has been estimated that roughly 2% to 7% of cases presenting with symptoms typical of ACS are instead suffering from TTC, but the cause of Takotsubo cardiomyopathy remains unknown [149,152]. It is much more common in post-menopausal women than other parts of the population and often apparently triggered by a stressful event. In approximately one-third of the cases, the onset is preceded by an emotially stressful situation. Roughly as often, the stressor is physical, related to e.g. exercise or a medical procedure. In the remaining one-third of the cases, no apparent trigger can be readily identified. [153-155]

An influential article published in 2005 reported that acute TTC associated with elevated circulating cathecholamine levels [156]. This study was, however, based on only 19 TTC cases, and its results have not been robustly replicated. Regardless, the catecholamine hypothesis remains perhaps the most commonly accepted model for the development of TTC, although a wide range of alternative explanations have been put forward [149].

The possible heritability of TTC is very hard to estimate. Being a rare, late-onset disease, both twin studies and studies based on pedigrees are difficult to conduct due to practicalities. Some have suggested that TTC might have an important genetic contribution, mostly based on isolated observations of multiple episodes of TTC in single individuals and multiple cases of TTC in some families [149,155,157-159]. Two candidate gene studies have reported an association of a SNP in the G-protein coupled receptor

kinase 5 with TTC, but the strenght of the evidence is not particularly high [160,161]. A study using whole-exome sequencing did not detect plausible rare variants in a sample of 28 TTC cases [162]. Similarly, a recent GWA study failed to identify genome-wide significant loci in a sample of 96 TTC cases and 475 controls [163]. Thus, the possible heritability of TTC remains unknown.

1.8.5.4 Other causes of heart failure

Heart failure is sometimes related to exposure to external agents. Radiation and some toxins can directly damage the heart [4]. The heart itself can be infected by a virus leading to myocarditis (inflammation of the heart), which sometimes causes heart failure [164]. An infection by streptococcal bacteria in the throat can trigger a strong autoimmune reaction known as rheumatic fever, which can damage multiple organs with the valves of the heart being among the most easily damaged. The resulting rheumatic heart disease has been mostly eradicated in the wealthy countries by the timely use of antibiotics, but rheumatic heart disease remains a significant burden in developing countries. [165] Chagas disease, caused by an infection by the protozoan parasite *Trypanozoma cruzi,* is also a regionally important cause of heart failure in Latin America [166].

A minor but significant proportion of HF is attributable to monogenic forms caused by rare alleles with high penetrance [167,168]. Among these, mutations affecting the contractile machinery of cardiomyocytes causing dilated and/or hypertrophic cardiomyopathies are prominent [168]. Other monogenic diseases in heart failure comprise channelopathies with mutations in the ion channels of the cardiomyocyte membrane (sarcolemma), proteins important to the morphogenesis of the heart associated with congenital heart disease, and mutations in genes related to neuromuscular disorders. [167,168] A common theme of the monogenic aetiologies of heart failure is their relatively early onset [167].

To date, more than 50 genes have been implicated in monogenic cardiomyopathies. [167,168] Especially striking are the high contribution of *TTN* mutations to the prevalence of dilated cardiomyopathy and the class of

diseases resulting from mutations in the gene *LMNA*. Truncating mutations in *TTN*, encoding the giant myocardial protein titin, are found in approximately 1 out of every 4 individuals suffering from dilated cardiomyopathy. [169] Different mutations in *LMNA,* encoding not one but three different proteins of the nuclear membrane, cause a staggering array of 10 diverse diseases, including dilated cardiomyopathy [170].

1.8.6 PREVENTION AND TREATMENT

Most cases of heart failure could probably have been prevented. Interventions in dietary sodium and cholesterol intake alone could decrease the burden of heart failure significantly, not only in the rich western countries, but also globally. Both primary and secondary prevention have already been succesful in some cases, such as the North Karelia Project in eastern Finland [171]. However, much more could still be gained in primary prevention focusing on the common, modifiable risk factors of cardiovascular disease.

The treatment of both acute and chronic heart failure partly depends on its immediate cause but is often similar regardless of the aetiology. The tretment is aimed at alleviating the symptoms and improving the prognosis of HF patients. Chronic heart failure with reduced EF is treated with angiotensin-converting enzyme (ACE) inhibitors, angiotensin receptor blockers, beta blockers, angiotensin receptor neprilysin inhibitors, and diuretic drugs, including mineralocorticoid receptor antagonists. In rare cases, if medical therapy is not alone sufficient, a cardioverter defibrillator or a cardiac resynchronization therapy device may be implanted or a heart transplant may be necessary. Treatment of chronic HF with preserved EF is not well established but can be targeted to alleviate the symptoms and increase well-being. Patients with arrythmias, e.g. atrial fibrillation, are treated with anticoagulants to reduce the risk of thromboembolic events, which is sometimes combined with digoxin to control heart rate. [4]

2 AIMS OF THE STUDY

The present study aimed at identifying and characterizing genetic variants associated with traits relevant to the development and diagnosis of heart failure in the Finnish population. Specifically, the studies comprising this thesis addressed the following traits:

1) Blood pressure (I, III)

2) Acute coronary syndromes (II)

3) Atrial and B-type natriuretic peptides (III)

4) Takotsubo cardiomyopathy (IV)

3 MATERIALS AND METHODS

3.1 STUDY SAMPLES

3.1.1 CHARACTERISTICS AND SAMPLING

The studies comprising this thesis analyzed data from several samples of the Finnish population, previously collected for epidemiological studies. Some have been designed and collected with the explicit aim of studying cardiovascular disease and its risk factors (COROGENE, TACOS, YFS), while others are more general surveys of the Finnish population (Health2000, FINRISK, NFBC, HBCS). Additionally, we used gene expression data from a cohort of autopsy donors collected in the USA [172]. Some key characteristics of the samples and cohorts are shown in Table 3.

The national FINRISK study collected cohorts of Finns every five years between 1972-2012 to monitor the risk factors of chronic non-communicable disease. The FINRISK cohorts analyzed in the present work (from the years 1992 to 2012) have been recruited from the population aged 25–64 years in five different geographical regions of Finland. In each region, the sampling was stratified for sex and 10-year age groups and random within each stratum. Together, the FINRISK cohorts form the most comprehensive epidemiological survey of Finns. [173,174]

The Health2000 survey, similar to the FINRISK cohorts, was collected by the National Institute for Health and Welfare to monitor the health and well-being of Finns. The study recruited adult (over 18 years old) participants from the 15 most populous cities and 65 randomly selected smaller regions. A subset of the cohort called GenMetS, including participants meeting the criteria for metabolic syndrome and their matched healthy controls, was analyzed in the present studies. [175,176]

The Helsinki Birth Cohort Study (HBCS) and the Northern Finland Birth Cohort (NFBC) study each recruited participants from a specific region in Finland. The HBCS included participants in Helsinki during 1934 to 1944 and were still living in Helsinki when the cohort was recruited in 1971 [177].

The NFBC66 cohort recruited the children of women with expected deliveries in 1966 in northern Finland (the provinces of Oulu and Lapland) with a follow-up data collection when the children had reached 1, 14, and 31 years of age [178].

The Cardiovascular Risk in Young Finns Study (YFS) was initiated to study the risk factors or cardiovascular diseases in the younger part of the Finnish population. In 1980, it recuited a cross-sectional sample of participants aged 3, 6, 9, 12, 15, and 18 years in in Helsinki, Kuopio, Oulu, Tampere, Turku, and their surroundings. While there have been several follow-ups of the participants, the data for the present study is from the follow-up made in 2001.

Three of the study samples analyzed collected cardiovascular disease cases. The COROGENE study recruited consecutive participants assigned to a coronary angiogram in the Helsinki University Central Hospital (HUCH) during 2006 to 2008 [179]. Of these, those diagnosed with MI or TTC were included in the data analyzed in this thesis. The Tampere Acute Coronary Syndrome Study (TACOS), recruited a consecutive series of participants admitted to the Tampere University Hospital and diagnosed with ACS during 2002-2003 [180].

In study IV, in addition to the COROGENE study participants, we included a sample of TTC cases, collected by the COROGENE investigators. This sample included retrospective cases recruited using the HUCH cardiology registry, and four prospective cases enrolled in 2010 [181]. For the additional sample of TTC cases, children recruited to the population-based CHILD-SLEEP study were used as controls and genotyped together with the additional TTC cases [182]. As the study data was from different sources, we matched the cases and controls based on genotyping array, the first two genomic principal components, and sex, to create the final study sample.

Table 3. Study samples.

Study	Cohort / year	Size*	Age	Sampling
The National FINRISK Study	1992,1997, 2002,2007	21942	25-74	Population-based from 5 geograpchical regions, stratified by 10-year age group and sex
FINRISK 2007 DILGOM	2007	518	25-74	FINRISK 2007 DILGOM extension study participants from the Helsinki/Vantaa region
The Health 2000 Survey	GenMetS	1997	30-75	Prevalent metabolic syndrome cases and controls from the population-based cohort
The Cardiovascular Risk In Young Finns Study (YFS)	1980 (2001)	2443	30-45	Population-based from Helsinki, Kuopio, Oulu, Tampere, and Turku
The COROGENE Study	2006-2008	2234	31-94	Consecutive ACS or TTC cases
TTC case series	2010	42	31-84	TTC cases recruited from the Helsinki region
CHILD-SLEEP	2011-2013	69	0-3	Infants born in Tampere University Hospital during 2011-2013
The Tampere Acute Coronary Syndrome Study (TACOS)	2002-2003	562	19-97	Consecutive cases with STEMI or NSTEMI (1/2002-3/2003) (or UA, 9/2002-3/2003)
The Northern Finland Birth Cohort (NFBC)	1996	5363	31	Birth cohort from the provinces of Oulu and Lapland
The Helsinki Birth Cohort Study (HBCS)	1934-1944	1619	57-70	Population-based of people born in Helsinki
The Genotype-Tissue Expression Project (GTEx)	2010-2015	190	21-70	Autopsy donors from USA

ACS, acute coronary syndrome; STEMI ST-elevation myocardial infarction; NSTEMI, non-ST elevation myocardial infarction; UA, unstable angina pectoris

* The sizes are reported as representative of the number of participants analyzed in the present studies and may be smaller than the total number of participants recruited.

3.1.2 GENOTYPING

With one exception (TACOS), the study samples have been genotyped using various genome-wide genotyping microarrays in multiple laboratories (Table 4). The TACOS study, used as a replication cohort in the study II, was genotyped for selected SNPs using the Sequenom iPLEX platform (Sequenom, San Diego, CA, USA).

The genotyping microarrays for the present studies have been supplied by Illumina (San Diego, CA, USA) and Affymetrix (Santa Clara, CA, USA, acquired in 2016 by Thermo Fisher Scientific, Waltham, MA, USA). The arrays are based on the hybridisation of target DNA to oligonucleotide probes and subsequent fluorescent detection of the hybridised DNA [183]. Although there are differences in the technological details between the manufacturers and arrays, they produce comparable results which are routinely combined in GWA studies using suitable statistical methods (e.g. meta-analysis).

3.1.3 ETHICS STATEMENT

All of the studies were conducted according to the principles expressed in the Declaration of Helsinki with the participants or their custodians giving written informed consent. The study samples have been reviewed and approved by their respective ethical committees.

Table 4. Genome-wide genotyping arrays.

Array	N markers*	DG	92	97	02	07	HB	NF	YF	H2	CG	TT	CS
Affymetrix Genome-Wide Human SNP Array 6.0	906 600		x	x	x								
Illumina HumanOmniExpress 12v1	733 202			x	x	x						x	x
Illumina HumanCoreExome-24	550 601		x	x	x	x							
Illumina HumanHap 610-Quad SNP Array	616 794	x	x	x	x	x				x	x		
Illumina HumanCNV370-Duo	370 404							x					
Illumina HumanHap 670k Custom	>600 000						x		x				

DG, DILGOM; 92 to 07, FINRISK 1992-2007 cohorts; HB, the Helsinki Birth Cohort Study; NF, the Northern Finland Birth Cohort 1966; YF, the Cardiovascular Risk in Young Finns Study; H2, the Health 2000 Survey GenMetS subsample; CG, the COROGENE study; TT, the Takotsubo cardiomyopathy case series; CS, the CHILD-SLEEP study

* the number (N) of markers as reported by the manufacturer

3.1.4 STUDY PHENOTYPES

3.1.4.1 Blood pressure

Systolic and diastolic blood pressure (SBP and DBP) were measured in each study cohort by the cohorts' investigators. Hypertension was defined as having at least one of the following: SBP > 140 mmHg, DBP > 90 mmHg, or known use of antihypertensive medication. In this study, mean arterial pressure and pulse pressure were defined as the arithmetic average and the difference of SBP and DBP, respectively.

3.1.4.2 Acute coronary syndromes

Diagnoses of ACS in study II were from different sources, namely the COROGENE investigators, the TACOS investigators, and from the Finnish hospital discharge and cause of death registries. In the COROGENE study, ACS was defined as as episode of typical acute chest pain with ≥ 50% stenosis in at least one coronary artery verified with coronary angiography. Those with ACS and elevated cardiac biomarkers were diagnosed with STEMI if they had persistent ST-segment elevations or with NSTEMI if they had either ST-depression or T-inversions. Additionally, those with typical symptoms and large cardiac enzyme release (creatine kinase-MBmass > 50 mg/l), but no significant coronary obstruction, were diagnosed with STEMI. ACS cases without elevated cardiac biomarkers were diagnosed with UA.

In TACOS, the diagnostic criteria were similar to COROGENE, but coronary stenosis was not used as a criterium as it was not investigated for a significant proportion of the study participants. Cases with symptoms or signs typical of ACS and elevated blood troponin (cTnI > 0.2 μg/L) were diagnosed with either STEMI or NSTEMI based on the study participants' ECG findings. Those presenting without elevated blood troponin were diagnosed with UA.

For the FINRISK study participants, diagnoses of ACS were retrieved from the hospital discharge and causes-of-death registries. They are thus

based on routine clinical work and potentially more heterogeneous than the diagnoses from the well-defined investigator-led studies. However, published validation studies have shown that they are of good quality [184].

3.1.4.3 Takotsubo cardiomyopathy

Takotsubo cardiomyopathy was diagnosed by the COROGENE investigators based on a set of criteria known as the "modified Mayo Clinic criteria". All of the following conditions were required for a diagnosis with TTC [181]:

1. New electrocardiographic (ECG) change (ST-elevation or T-wave inversion or both) or cardiac-enzyme release (troponin or creatine kinase) or all of these
2. Transient hypo-, dys-, or akinesis of the left ventricular midsection with or without involvement of the apex extending over a single epicardial coronary vascular bed
3. Absence of obstructive coronary artery disease (coronary luminal narrowing less than 50%) or plaque rupture confirmed by coronary angiogram (CAG)
4. Absence of myocarditis or pheochromocytoma

3.1.4.4 Natriuretic peptides

Natriuretic peptide measurements for the study III were made in the MORGAM Biomarker Laboratory (University of Mainz, Germany) using three immunoassays: C-terminal B-type natriuretic peptide was measured using the Abbott Architect i2000 BNP assay (Lake Bluff, IL, USA). N-terminal proBNP was measured using the Roche Elecsys 2010 proBNP assay (Roche Diagnostics, Indianapolis, IN, USA). Mid-regional proatrial natriuretic peptide was measured using the B.R.A.H.M.S. MR-proANP KRYPTOR assay (B·R·A·H·M·S Aktiengesellschaft, Hennigsdorf, Germany). [185]

All of the assays used are commercially available. Regrettably, the manufacturers are not entirely helpful in sharing the specifics of the assays'

principles. The online table "Analytical characteristics of commercial MR-proANP, BNP and NT-proBNP assays as per the manufacturer" maintained by The International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) has proven to be a valuable resource in this regard [186]. While the natriuretic peptide assays have slightly distinct operating principles, all are immunoassays utilizing a capture antibody to bind the target molecule into place (with high affinity) and a detection antibody to measure its quantity in the sample (with high specificity) using optical detection [185,186]. See Figure 8 for a schematic representation of the natriuretic peptide regions targeted by the assays.

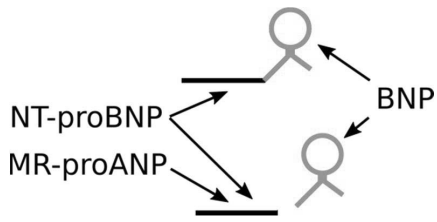

Figure 8. Natriuretic peptide assays. The NT-proBNP and MT-proBNP assays bind to the N-terminal region of the prohormone, whereas the BNP assay binds the C-terminal region containing the mature hormone. As a significant proportion of circulating proBNP is uncleaved, the NT-proBNP and BNP assays bind to either the cleaved N-terminal fragment and the mature C-terminal hormone, or the intact uncleaved prohormone. ProANP is cleaved prior to secretion and the MR-proANP assay is thought to measure the concentration of the cleaved N-terminal fragment.

3.1.4.5 Gene expression

In studies II and III, we investigated the correlation between genotypes and gene expression levels. In study II, we used gene expression measurements from the DILGOM study, made using the Illumina HT-12 array and total RNA extracted from whole-blood samples as previously described [187,188]. The arrays had been background corrected and normalized so that the signal intensity distributions for all samples on all arrays were equal. All participants had been measured using two technical replicates, and nine participants had been excluded based on poor correlation between the replicates.

In study III, we used gene expression data from the Genotype-Tissue Expression (GTEx) project [172,189]. GTEx has profiled gene expression

from a multitude of tissues from deceased donors, using RNA sequencing with a library construction protocol (Illumina TruSeq) that uses polyA selection. For study III, we used the GTEx eQTL analysis data from left ventricular and atrial appendage tissue samples (release V6, Oct 6th 2016) [189].

3.2 STATISTICAL METHODS

3.2.1 GENOTYPE IMPUTATION

Genotype imputation was used in all of the studies comprising this thesis. Two different algorithms, MACH and IMPUTE2, were used in combination with different reference panels [190,191]. Both MACH and IMPUTE2 algorithms use a Markov chain Monte Carlo (MCMC) model to impute missing genotypes based on a set of phased reference haplotypes. A separate pre-phasing step prior to imputation with IMPUTE2 was done with the Shape-IT algorithm for a part of the data to speed up the imputation process, as has been recommended by the IMPUTE2 authors [192].

Althogether, seven different imputation reference panels were used, reflecting the progression of the imputation methods. The reference datasets used for all studies contained at least some Finnish participants, with the exception of the HapMap 2 panel used for a part of the data in the first study. The reference panels are listed in Table 5.

Table 5. Imputation reference panels.

Panel	N	N FIN*	N variants	Studies
HapMap 2 CEU [193]	90 (30 trios)	0	3 204 709	I
1000G Pilot CEU (8/2009) [194]	60	0	7 724 854	I
HapMap 3 [195]	1 397	0	1 481 135	I
HapMap 3 Finnish [196]	81	81	1 163 280	I
1000G Phase 1 (3/2012) [197]	1 092	93	39 706 715	II,III
1000G Phase 1 (9/2013) [197]	1 092	93	37 878 821	II,III,IV
Custom Finnish [198]	1 941	1 941	13 625 231	III

1000G, the 1000 Genomes Project

* N FIN: the number of Finnish samples in the panel

3.2.2 IMPUTATION AND TRANSFORMATIONS OF PHENOTYPIC DATA

Study participants with missing phenotype values were excluded in all analyses, with the exception of the GWAS for natriuretic peptide traits in study III. In study III, missing values for BNP and NT-proBNP were imputed using multiple imputation methods in the "mice" package of R with subsequent pooling of the results [199,200]. Similar to conventional imputation resulting in a dataset with the missing values "filled in", multiple imputation produces multiple imputed versions. It has been reported to be a robust way to deal with the various potential issues arising from missing data, particularly when the rate of missingness is large [200]. In retrospect, as the rate of missing data in study III was small, multiple imputation was perhaps an excessive precaution.

The derivation of P-values for standard linear regression assumes the regression residuals are normally distributed. We thus transformed the quantitative traits analyzed to achieve normality. For systolic blood pressure and the gene expression traits (probe intensities), log-transforming the data resulted in an approximately normal distribution. For the natriuretic peptide traits in study III, we ensured normality of the traits by using inverse normal transformation, which first ranks the observations and then sets the values of the ranked observations so that they are distributed according to the standard normal distribution [201].

3.2.3 ASSOCIATION TESTS

We used linear regression as the test of association for quantitative traits (blood pressure, gene expression, natriuretic peptides) in studies I to III. In study III, we also used a method based on LASSO regression to fine-map the loci where we detected a genome-wide significant association.

For the binary disease endpoints in substudies II and III, we used logistic regression (for prevalent cases) or the Cox proportional hazards model

(incident ACS cases in study II). In study IV, where the participants were matched case-control strata, we used conditional logistic regression to take into account the matching of the study participants into the multiple strata. The software packages implementing the tests used in the studies are listed in Table 6.

In substudies I to III, we modeled the genotypes using an addititive genetic model with the reference homozygotes, heterozygotes, and the coded homozygotes modeled using 0, 1, 2 coding, respectively. If the variants had been imputed with uncertainty, we used the expected genotype counts, also known as "genotype dosages", to take into account the probabilistic nature of the data. Additionally, in study IV, we made separate recessive and dominant models, where either the reference homozygote only (for the dominant model) or the reference homozygotes and the heterozygotes (for the recessive model) were coded as 0 and the other genotypes as 1.

In substudies II and III, we wanted to assess whether variants associated with a given phenotype were also associated with the expression of nearby genes. In study II, this was limited to tabulating and plotting the associations and noting whether they passed a nominal significance limit. In study III, we developed a more robust approach based on the correlation of variant P-values and effect sizes. We first refined the set of variants to be included in the comparison by requiring they are within the genomic region showing an association to the trait, excluding variants with very different LD patterns in the samples used for the GWA study and the gene expression tests. We then investigated the Pearson's correlation between the effect estimates (which do not depend on allele frequencies) and Spearman's rank correlation for P-values (which do depend on allele frequencies), deriving the statistical significance of the comparisons empirically. This approach is documented in detail in the Supplementary Information for study III.

Table 6. Genetic association tests.

Study	Trait	Test	Software (Test)	Software (Meta-analysis)
I	SBP	LIN-R	ProbABEL, SNPTEST	MetABEL, METAv1.2
	DBP			
	PP			
	MAP			
II	ACS	LOG-R, COX	PLINK, SNPTEST, R survival	GWAMA v2.1
	Gene Exp.	LIN-R	SNPTEST	-
III	Hypertension	LOG-R	R glm	R metafor
	SBP	LIN-R		
	DBP			
	BNP	LIN-R, LASSO	SNPTEST, LLARMA, R glm	R metafor
	BNP:NT-proBNP			
	NT-proBNP			
	MR-proANP			
IV	TTC	C-LOG-R	R clogit	R metafor

ProbABEL [202]; MetABEL [202]; SNPTEST [203]; METAv1.2 [204]; PLINK [205]; GWAMA [206]; R glm [207]; R metafor [208]; LLARMA [209]; R clogit [210]

C-LOG-R, conditional logistic regression; LIN-R, linear regression; LOG-R, logistic regression; LASSO, LASSO regression

3.2.3.1 Regression models

We included several covariates in the regression models in order to attenuate any possible confounding and reduce statistical noise in the analysis. All of the primary analyses were adjusted for the sex of the study participants. We selected additional covariates based on prior information of their importance or based on exploratory analyses in the study showing a significant association ($P<0.05$) with the analysed trait. Not all of the covariates were available or relevant for all datasets, so we adapted the specific selection of covariates specifically to each data source. The variables included in the models are listed in Table 7.

Table 7. Regression models.

Study	Trait	Covariates
I	Blood pressure	age, sex, smoking habits, alcohol consumption, body mass index
II	MI (GWAS)	age, sex, genomic principal components
II	MI (TACOS)	-
II	Incident MI (FINRISK)	age, sex, genomic principal components, study year, geographical region, genotyping batch, systolic blood pressure, blood pressure medication, total cholesterol, HDL-cholesterol, current smoking, prevalent diabetes
II	Gene expression	none
III	Natriuretic peptides	age, age squared, sex, geographical sampling region, body mass index, current smoking, systolic blood pressure, glomerular filtration rate estimated using cystatin C and creatinine as proxies (eGFR), genotyping batch
III	Blood pressure	age, sex, body-mass index, genomic principal components, current blood pressure medication use, study year, genotyping batch
IV	TTC	sex, genotyping batch, genomic principal components

MI, myocardial infarction; GWAS, genome-wide association study; eGFR, estimated glomerular filtration rate; HDL, high-density lipoprotein; TTC, Takotsubo cardiomyopathy

3.2.4 HERITABILITY ESTIMATES

We used the GCTA (Genome-wide Complex Trait Analysis) software package in study III to estimate the proportion of variance in the natriuretic peptide traits explained by the genome-wide genotype data [211]. This method is based on first estimating the genetic relatedness between all study participants and then fitting a model to the data which expresses the observed trait values as a function of the genetic relatedness estimates. The heritability estimates calculated with GCTA are thus based on genetic relatedness across the specific set of variants used in the calculations. Furthermore, the estimates are based on an additive model and thus are comparable to the "narrow sense" h^2 heritability metric. The method is analogous to estimating heritabilities using the twin study design or pedigrees where the genome-wide genetic relatedness is given by the pedigrees and does not need to be measured.

3.2.5 ESTIMATES OF STATISTICAL POWER

In study IV, we estimated the statistical power of our study sample using data from the Finnish participants of the 1000 Genomes Project, using $P=5 \times 10^{-8}$ as the limit for genome-wide statistical significance. To estimate how well our study genotypes correlated with possible causal variants, we sampled 1,000 variants from each MAF percentile and treated them as causal alleles. We identified the best tag SNPs in our data in terms of their r^2 with the causal alleles, only considering directly genotyped SNPs and SNPs with high imputation accuracy (info > 0.95). We varied the disease models (multiplicative, recessive, dominant) and the genotype relative risks (GRR, from 1 to 10, increments of 0.25 units) of the causal alleles and, across the 1,000 variants in each MAF percentile, calculated the average power to detect the association for each GRR, using the best tag SNPs. For the multiplicative disease model, we used the "GPC" function of the GeneticsDesign R package. For the dominant and recessive disease models, we used the "2p2n.test" function from the pwr package for R. [207] We used 1/5000 as the unknown disease prevalence parameter and used the the allele frequency and LD parameter values obtained from the 1000 Genomes Project data. We multiplied the r^2 coefficient by the "info" metric from the imputed analysis dataset to take into account uncertainty in the imputation.

4 RESULTS AND THEIR EVALUATION

4.1 STUDY I: BLOOD PRESSURE

Elevated blood pressure increases the risk for many complex cardiovascular diseases. Such a link has been suspected also for intracranial aneurysms (IA), rupture-prone pouches in cerebral arteries. Diagnosis of IA, however, requires significant technological resources (computerized tomography, magnetic resonance imaging, or cerebral angiography), leading to a lack of large epidemilogical studies needed to conclusively establish or reject a link between elevated blood pressure and IA. In study I, we used the suspected link between elevated blood pressure and IA as a working hypothesis, predicting that some of the genetic variants associated with IA are in fact related to blood pressure. We, therefore, aimed at identifying variants associating with both blood pressure and IA. The successful identification of such variants would strenghten the hypothesis and, hopefully, lead to the identification of new blood pressure and IA susceptibility loci. [212]

We selected variants which had shown at least suggestive association with IA in a recent multinational GWAS (posterior probability for association > 0.1 as estimated by the authors of the IA GWAS) [213]. Altogether, 41 variants from 19 distinct genomic regions passed this threshold and were included in our study. We used the Health2000 GenMetS study samples as an initial discovery cohort, with a liberal P-value limit ($P<0.1$) to prune out variants lacking any association with the blood pressure traits of our study (SBP, DBP, PP, MAP). We then followed the remaining variants in the YFS, HBCS, and NFBC1966 cohorts. Finally, we queried the association of our most statistically significant result in data from the International Consortium for Blood Pressure Genome-wide Association Studies (ICBP-GWAS).

SNPs in two of the 19 genomic regions on chromosomes 2 and 5 passed the liberal P-value threshold in H2000. Of these, only the three SNPs (rs570682, rs2287696, and rs335206) on chromosome 5 were robustly associated with blood pressure across the Finnish cohorts (Table 8; Figure 9). These SNPs were also associated with blood pressure in the data from the

multinational ICBP-GWAS consortium. Overall, the association of the SNPs was more statistically significant for SBP than for the other traits tested. The statistically most significant association was with rs2287696 and SBP ($P=8.13 \times 10^{-7}$ in the meta-analysis of the Finns and ICBP-GWAS). The three SNPs are in strong LD, not only in Finns but also globally, with pairwise D' > 0.8 across the 1,000 Genomes Project Phase 3 data [6]. Thus, the associations likely represent a single signal tagged by all three SNPs, rather than multiple underlying causal variants.

The three above-mentioned SNPs are located in one of the introns of the PR/SET domain 6 gene *PRDM6*. To inspect this region in detail, we studied the SBP association of all SNPs present in 1000 Genomes imputed data from the four Finnish cohorts. The region of association covered the entire *PRDM6* gene and extended some 150 kbp downstream from its last exon.

Both the location of the BP-associated SNPs and published studies on the function of PRDM6 led us to attribute the blood pressure association to this gene. PRDM6 has been reported to mediate the transition of vascular smooth muscle cells (SMC) from a proliferative to a differentiating phenotype, inhibiting differentiation and promoting proliferation [214]. This transition is highly relevant to both hypertension and IAs. Thus, we speculated in the original publication that the associations we detected stem from a regulatory effect on *PRDM6*, increasing its expression and SMC proliferation in the arterial wall. Unfortunately, at the time of the original study, we did not have access to high-quality data to confirm of refute this hypothesis.

Recent data from the GTEx consortium, published in 2017, lend at least some support to our initial interpretation of the results. The A allele of rs2287696, correlated with increased systolic blood pressure and IA risk, is indeed highly significantly associated with the expression of *PRDM6* in samples from the tibial artery ($P=4.9 \times 10^{-7}$) and aorta ($P=2.4 \times 10^{-5}$) (Figure 10) [189]. However, the direction of the association is opposite to what we predicted in the original publication: The rs2287696 A allele is associated with decreased, rather than increased, *PRDM6* expression. Therefore, based on this data and the report of PRDM6 promoting SMC proliferation, the A allele should associate with decreased SMC proliferation.

Table 8. Association of IA SNPs with systolic blood pressure in the Finnish cohorts.

SNP	Chr	BP	Alleles	MAF	H2000*	YFS*	NFBC1966*	HBCS*	P_{META}	$beta_{META}$*
rs1429412	2	198190894	A/G	0.5	0.75 (0.64)	-0.27 (0.42)	-0.15 (0.24)	0.05 (0.85)	0.665	-0.08 (0.19)
rs12472355	2	198205840	C/A	0.5	0.73 (0.64)	-0.24 (0.42)	-0.11 (0.24)	-0.03 (0.85)	0.771	-0.06 (0.19)
rs787997	2	198216271	A/G	0.4	0.62 (0.65)	-0.11 (0.42)	-0.08 (0.24)	0.25 (0.84)	0.966	-0.01 (0.19)
rs787994	2	198223121	T/C	0.4	0.68 (0.65)	0.03 (0.42)	-0.03 (0.24)	0.26 (0.85)	0.753	0.06 (0.19)
rs570682	5	122449650	T/C	0.2	1.48 (0.69)	1.22 (0.47)	0.71 (0.27)	0.46 (0.98)	4.80×10^{-5}	0.87 (0.22)
rs2287696	5	122460332	G/A	0.2	1.68 (0.71)	1.18 (0.49)	0.67 (0.28)	0.93 (1.03)	6.81×10^{-5}	0.89 (0.22)
rs335206	5	122504566	T/C	0.4	1.02 (0.60)	0.85 (0.40)	0.74 (0.24)	0.60 (0.84)	3.01×10^{-5}	0.79 (0.19)

Genomic coordinates are reported using the GRCh37 reference genome version. Variants associated with increased risk for IA were set as the coded or effect allele and are underlined. The effect estimates are reported from linear regression models adjusted for gender, age, BMI, smoking status, and alcohol use. P_{META} and $beta_{META}$ from inverse variance weighted fixed effect meta-analysis combining the statistics across all cohorts.

Chr, chromosome; BP, base pair; MAF, minor allele frequency; H2000, the Health 2000 survey; YFS, the Cardiovascular Risk In Young Finns Study; NFBC1966, the Northern Finland Birth Cohort 1966; HBCS, the Helsinki Birth Cohort Study

* Statistics reported as beta or regression coefficient (standard error)

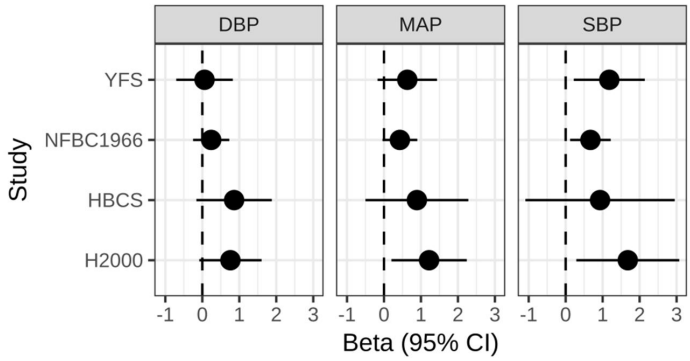

Figure 9. Association of rs2287696 with blood pressure in the four Finnish cohorts. The regression coefficients or betas are shown in mmHg per allele.

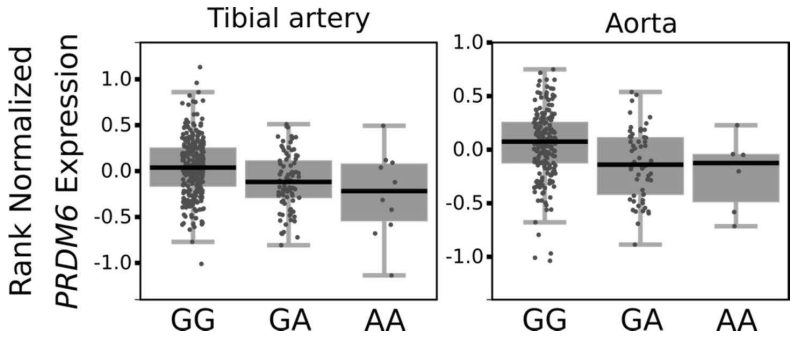

Figure 10. Rs2287696 and *PRDM6* expression in arterial tissue samples from the GTEx consortium for the three genotype classes. Figures modified from those retrieved on 1.11.2017 from the GTEx web portal at https://www.gtexprotal.org.

4.2 STUDY II: ACUTE CORONARY SYNDROME

Myocardial injury in acute coronary syndromes is among the most common causes of heart failure. Elevation of the ST-segment in electrocardiography during acute myocardial infarction is thought to reflect the transmural character of myocardial injury. According to this model, STEMI is caused by a complete blockage of a coronary artery, while NSTEMI results from incomplete or transient blockage. The elevation of the ST-segment in electrocardiography thus reflects a transmural infarction (myocardial injury

spanning the entire thickness of the heart muscle) caused by complete blockage of a coronary artery. In contrast, non-ST segment elevation in MI reflects subendocardial injury, where the cells closest to the incomplete block in the coronary artery near the surface of the heart still receive enough oxygen to survive, but cells deeper within the myocardium suffer from hypoxia or anoxia. (Figure 7) [215,216]

In study II, we wanted to compare and indentify genetic variants conferring susceptibility to STEMI or NSTEMI [217]. Although many studies have been published on the genetics of CAD or MI in general, the division of MI into STEMI and NSTEMI has received little attention from a genetical point of view. To address this distinction, we employed GWA analysis in the COROGENE study sample, stratifying the participants based on whether they suffered from NSTEMI or STEMI.

The GWAS identified SNPs on chromosome 1 associating with NSTEMI showing much weaker association with STEMI (Figure 11, Table 9). Variants associated with NSTEMI cover a region containing the genes *DENND2D*, *CEPT1*, and *DRAM2*. The most statistically significantly associated variants were located in an intron of *DENND2D*, upstream relative to *DRAM2* and downstream relative to *CEPT1*.

To replicate the GWAS result, we tested the most statistically significant directly genotyped SNP rs656843 for association with STEMI and NSTEMI in participants of the TACOS study. Additionally, we also tested it for association with incident MI cases from the FINRISK cohorts. The association with NSTEMI replicated in TACOS. It was also similar in direction in the smaller set of incident cases from FINRISK, i.e. the the risk allele was more common in NSTEMI cases, but in FINRISK the association was not statistically significant (P=0.43).

The overall most statistically significant association was with NSTEMI when we combined the results from the case-control samples (COROGENE and TACOS) using meta-analysis (rs656843; odds ratio 1.57, $P=3.11 \times 10^{-10}$). We followed this result in non-Finnish populations by examining data from two large international GWAS meta-analyses of coronary artery disease, where rs656843 did not show any statistically significant association

(P>0.05). The locus was, therefore, genome-wide significant and replicated for NSTEMI in Finns but did not associate with CAD in two international meta-analyses.

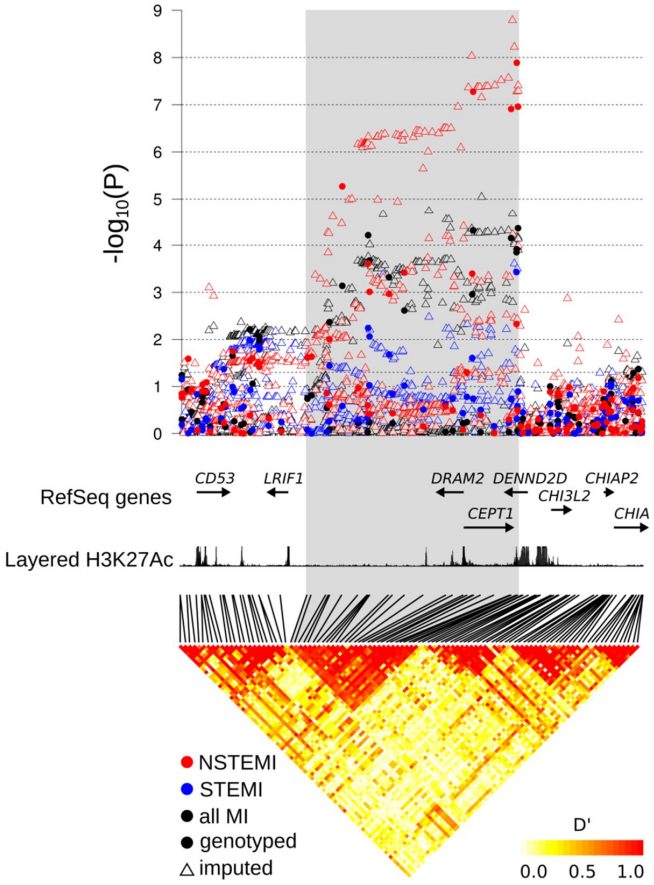

Figure 11. Association of genetic variants with MI on chromosome 1. Genes are depicted as annotated in the RefSeq database. Potential regulatory sites, indicated by histone 3 lysine 27 acetylation (Layered H3K27Ac), as reported by the Encode consortium, are shown under the genes [28,29]. Linkage disequilibrium is shown calculated from the directly genotyped SNPs of the GWAS sample.

Table 9. Association of rsr656843 with MI

Trait	Sample	MAF Controls (N)	MAF Cases (N)	OR* (95% CI)	P
NSTEMI	COROGENE	0.15 (1,576)	0.21 (962)	1.63 (1.38-1.93)	1.22×10^{-8}
	TACOS	0.14 (566)	0.19 (389)	1.44 (1.12-1.86)	0.0049
	Case-control combined†	0.15 (2,142)	0.20 (1,351)	1.57 (1.36-1.81)	3.11×10^{-10}
	FINRISK	0.17 (16,143)	0.18 (163)	1.13 (0.84-1.51)	0.43
STEMI	COROGENE	0.15 (1,576)	0.16 (614)	1.07 (0.88-1.29)	0.5
	TACOS	0.14 (566)	0.17 (173)	1.27 (0.91-1.77)	0.16
	Case-control combined†	0.15 (2,142)	0.16 (787)	1.11 (0.94-1.31)	0.2
	FINRISK	0.17 (16,143)	0.16 (99)	0.97 (0.65-1.43)	0.87
All MI	COROGENE	0.15 (1,576)	0.19 (1,579)	1.33 (1.15-1.54)	1.27×10^{-4}
	TACOS	0.14 (566)	0.18 (562)	1.38 (1.10-1.74)	0.0062
	Case-control combined†	0.15 (2,142)	0.19 (2,141)	1.34 (1.19-1.52)	2.58×10^{-6}
	FINRISK	0.17 (16,143)	0.16 (484)	1.00 (0.84-1.19)	0.98

MAF, minor allele frequency; OR, odds ratio; MI, myocardial infarction; STEMI, ST elevation myocardial infarction; NSTEMI, non-ST elevation myocardial infarction;

Case-control samples were analyzed using logistic regression including age, sex, and the first ten genomic principal components (COROGENE) or no covariates (TACOS) in the model. FINRISK was analyzed with the Cox proportional hazards model stratified by study year, geographical region, and genotyping batch with gender, systolic blood pressure, blood pressure medication, total cholesterol, HDL-cholesterol, smoking, and diabetes used as covariates.

* For FINRISK, the effect is reported as the hazard ratio

† Results from COROGENE and TACOS combined using inverse-variance weighted fixed effects meta-analysis

Unequal sample sizes in the GWAS would be the simplest explanation for observing a more statistically significant association for NSTEMI (N=962) than for STEMI (N=614). The association was, however, statistically less significant when analyzing both MI types together, despite the larger sample size in this analysis setting. To formally investigate whether the effects were different for NSTEMI and STEMI, we calculated posterior probabilities of statistical models where rs656843 conveyed either no risk, the same risk, a higher risk for NSTEMI, or risk only for NSTEMI. Out of these models, the latter two with risk specific to NSTEMI alone (posterior probability 0.71) or a related but smaller risk for STEMI (posterior probability 0.29) were the most probable. The effect underlying the association, therefore, appeared either specific to or at least stronger for NSTEMI.

Although other causes do exist, both NSTEMI and STEMI are, in general, mostly complications of CAD. Furthermore, the COROGENE study participants had been studied using coronary angiography, confirming CAD as the probable cause of the MI. The stronger association observed with NSTEMI, however, cannot simply be explained as arising from increased susceptibility to CAD, as such an effect should affect both STEMI and NSTEMI in a similar manner. To confirm this, we explored the associations of variants previously reported with GWAS for CAD or MI. A total of 50 such SNPs were present in our GWAS data. None of them showed statistically significant differences in allele frequencies between the STEMI and NSTEMI cases when the P-values were Bonferroni corrected for the 50 tests made. The NSTEMI association detected in Finns was, therefore, quite different than the previously identified CAD risk loci.

The most statistically significant difference in STEMI vs. NSTEMI for the previously published CAD or MI risk SNPs was with rs514659. Intriguingly, it was also the only SNP among the 50 specifically associated with MI in the presence of CAD rather than the risk for CAD [218]. The C allele of rs514659 was slightly more common in STEMI cases (C allele frequency in STEMI patients 48.06% vs. 43.26% in NSTEMI; P=0.0066). Rs514659 tags one of the genetic variants responsible for the ABO blood groups, the C allele

correlating with the O blood group. This data may suggest that the O blood group associates with larger risk for STEMI than for NSTEMI.

Typical of GWAS, the SNPs associating with NSTEMI in our study were not predicted to change the amino-acid sequence of proteins coded by the genes in the locus. This suggests that the NSTEMI association depended on an underlying effect on gene expression. To test this hypothesis, we used RNA expression data measured from whole blood in 513 healthy participants of the DILGOM study, an extension of FINRISK 2007. We tested the SNPs in the region for association with the expression of the nearby genes and identified two independent signals for the expression of *DRAM2*. The primary signal was much stronger but did not correlate with the association of the SNPs with NSTEMI. The secondary signal, detected when we set the most significant *DRAM2* expression SNP as a covariate, was statistically weaker but was somewhat correlated with the results for NSTEMI. These results are somewhat inconclusive, although they offer at least some support to the interpretation that the NSTEMI association may be related to the expression of *DRAM2*.

4.3 STUDY III: NATRIURETIC PEPTIDES

Atrial and B-type natriuretic peptides (ANP and BNP) are unique biomarkers of heart failure. Despite their prominence in cardiovascular physiology, including diagnostics and attempts at using them as therapeutic agents, surprisingly few studies have addressed their role in humans during approximately normal homeostatic conditions. [101,219] In study III, we investigated the relationship between genetic variation, natriuretic peptides, and blood pressure in the general population (Table 10). Specifically, we used GWA analysis to study the circulating concentrations of midregional proatrial natriuretic peptide (MR-proANP), BNP, and N-terminal proBNP (NT-proBPN). As proBNP is cleaved in the circulation after it has been secreted by cardiomyocytes, we also analysed the ratio of BNP to NT-proBNP concentrations as a proxy of their processing in the circulation [101].

We first estimated the SNP-based heritabilities of the traits. The point estimates of phenotypic variance explained by our entire genome-wide data ranged from 13.9% for MR-proANP to 23.0% for NT-proBNP (Table 11). Most of the variance was, therefore, explained by factors other than the genetic variation captured by our dataset, but all traits had a moderate genetic contribution as well. The precision of the estimates we obtained was coarse, preventing a meaningful comparison between the four traits.

We detected genome-wide significant associations in four genomic loci (Figure 12, Table 12). Replication of the associations was statistically significant for all but one of the loci (rs701041 on chromosome 12 near the nuclear receptor corepressor 2 gene *NCOR2*). Fine-mapping the successfully replicated loci identified five independent variants in the associated region on chromosome 1, two independent variants in the chromosome 12 locus, while a single signal explained the association on chromosome 8. Thus, a total of eight independent variants in three genomic regions associated with at least one of the natriuretic peptide traits of our study.

On chromosome 1, the natriuretic peptide associated SNPs were located in and around the adjacent genes *NPPA* and *NPPB*, which encode proANP and proBNP, respectively. Associations with natriuretic peptides have been reported in this region previously, but the prior studies did not have both high-resolution genomic coverage and mesurements for both ANP and BNP, and had been unable to specifically attribute the associations to either ANP or BNP [106,108-110,220]. In our study, we were able to divide the associatied variants into those associating with MR-proANP and those associating with BNP or NT-proBNP. None of the SNPs associated with both ANP and BNP, which is somewhat surprising taking into account that *NPPA* and *NPPB* are separated by less than 10 kbp.

Three of the five indepently associated SNPs on chromosome 1 associated with the circulating concentration of MR-proANP. They are located approximately 40 kb downstream from *NPPA*, within the transcribed regions of the genes *MTHFR* and *C1orf167*. We estimated that the effect sizes of these three SNPs were approximately 2.5 to 5.0 pmol/L. The median MR-proANP concentration of the GWAS samples was 41.3 pmol/L (interquartile range

25.2 pmol/L), so the effects of the SNPs were quite substantial compared to typical values in the study population. Comparing the association of SNPs with MR-proANP with their effect on gene expression in left ventricular and atrial appendage tissue samples from the GTEx consortium indicated that, as expected, the SNP effects for MR-proANP and the expression of *NPPA* were positively correlated (Figure 13). The SNPs detected as associated with MR-proANP are, therefore, likely related to regulatory variants affecting the expression of *NPPA* in the heart.

Two independent SNPs on chromosome 1 associated with BNP or NT-proBNP. Rs198379 associated with both BNP and NT-proBNP, with an estimated effect size of 4.5 pg/mL (for BNP) and 9.6 pg/mL (for NT-proBNP) per allele. The other SNP, rs61761991, was extremely significantly associated with NT-proBNP ($P=8.76 \times 10^{-68}$) and BNP:NT-proBNP ratio ($P=4.81 \times 10^{-103}$)[2] but did not associate with C-terminal BNP. The two SNPs had thus very different association profiles. Rs198379 showed, as expexted, an association to measurements both from the C-terminal and N-terminal ends of proBNP. The association of SNPs with BNP was also correlated with their association to expression of *NPPB* in the heart, indicating the presence of an underlying causal variant affecting *NPPB* expression (Figure 13). Rs61761991, however, only associated with the N-terminal measurement.

Rs61761991 is an amino-acid changing SNP located in the N-terminal proBNP prohormone (NP_002512.1:p.Arg72His). This region is used as the antigen to prepare the capture antibody for the NT-proBNP assay employed in our study (and apparently most other NT-proBNP assays as well) [186]. Rs61761991 appeared to effectively block the signal of the NT-proBNP assay, as those heterozygous for this SNP had roughly 50% smaller NT-proBNP values on average than the common homozygotes (Figure 14). The rarer T allele of rs61761991 is practically absent in most populations but enriched in Finns where the its frequency is roughly 3% [221]. Assuming Hardy-Weinberg equilibrium, we can thus estimate that approximately one in twenty Finns are rs61761991 heterozygotes and should present a measured

[2] 4.81×10^{-103} = 0.00000 00481

concentration of NT-proBNP some 50% lower than extrapolated from the corresponding C-terminal BNP value. One in every thousand Finns is expected to be a TT homozygote with very small measured NT-proBNP concentration, regardless of the circulating BNP concentration.

SNPs on chromosome 8 associated with BNP:NT-proBNP ratio and NT-proBNP, but not with MR-proANP or BNP. Of these SNPs, the most statistically significant SNP in the GWAS was rs7000551. The SNP is located in an intron of solute carrier family 39, member 14 gene *SLC39A14*, approximately 20 kbp upstream from the protein phosphatase 3 catalytic subunit gamma gene *PPP3CC*. We did not detect statistically independent secondary associations in this locus, suggesting that the signal we detected relates to a single causal allele correlated with rs7000551. Gene expression data from GTEx indicated that the BNP:NT-proBNP associated variants on chromosome 8 also correlated with the expression of *PPP3CC* in both ventricular and atrial tissue samples (Figure 13). Alleles associated with increased BNP:NT-proBNP ratio correlated with reduced expression of *PPP3CC*.

Protein phosphatase 3, also known as calcineurin, has been shown to be an important regulator of cardiac hypertrophic signaling [222]. It is an enzyme consisting of three subunits: a catalytic subunit, a regulatory subunit, and a subunit known as calcineurin A. In humans, there are three alternative catalytic subunits (alpha, beta, gamma), with *PPP3CC* encoding the gamma catalytic subunit. Differences in *PPP3CC* expression should, therefore, be reflected as differences in the relative abundancies of calcineurin isoforms. This data indicates that the association of SNPs on chromosome 8 near *PPP3CC* with the BNP:NT-proBNP ratio may relate to differences in *PPP3CC* expression in the myocardium.

The third genome-wide significant locus we detected was located on chromosome 12. A prior study had reported an association with an SNP in this region with NT-proBNP [107]. In our data, we identified two independent SNPs in this locus, both associated with NT-proBNP and the BNP:NT-proBNP ratio. Similar to the chromosome 8 locus near *PPP3CC*, the chromosome 12 SNPs did not associate with BNP or with MR-proANP.

Comparison with gene expression failed to highlight any of the genes in the region as likely candidates. The authors of the previous study had pointed to the polypeptide N-acetylgalactosaminyltransferase 4 gene *GALNT4* located in the NT-proBNP associated region as a potential culprit, as it initiates O-linked glycosylation and proBNP is known to be glycosylated [107]. Our observation of no association with BNP offers some support to this idea, as the C-terminal region containing the active hormone is not glycosylated.

Natriuretic peptides are thought to be important regulators of blood pressure. We explored this by testing the genetic variants identified with natriuretic peptide GWAS for association with blood pressure in multiple population-based cohorts of Finns, identifying a weak but consistent association with BP for the MR-proANP SNPs. The SNPs associated with BNP, NT-proBNP, or their ratio did not associate with blood pressure in our data. The MR-proANP increasing SNP alleles associated with approximately 0.25 mmHg (diastolic BP) and 0.50 mmHg (systolic BP) differences in blood pressure.

To assess the combined effect of the three MR-proANP associated SNPs on blood pressure, we formed a summary variable summing the number of MR-proANP increasing alleles for all three SNPs (thus ranging from 0 to 6). This variable explained 2.36 % of the variance in MR-proANP concentration and associated with a 9 % decrease in the odds ratio for hypertension (OR=0.91; SE=0.0283; $P=8.2 \times 10^{-4}$). We also estimated the fraction of hypertension cases attributable to the three-SNP allele counting variable as 11.8%, surprisingly high given the moderate effects of the individual SNPs on blood pressure. Thus, a moderate decrease in MR-proANP was estimated to substantially increase the prevalence of hypertension in our study samples.

Table 10. Sample characteristics in the natriuretic peptide study.

	GWAS Sample*	Replication Sample*	Blood Pressure Study Sample†
N	4,932	1,373	27,059
Age (years)	46.18 (21.39)	48.67 (20.5)	42.41 (25.42)
Females (n/%)	2,592 (52.55%)	711 (51.78%)	14,377 (53.13%)
Body Mass Index (kg/m2)	25.73 (5.473)	25.97 (5.116)	25.69 (5.714)
Diastolic Blood Pressure (mm Hg)	82 (15)	82 (14)	80 (16)
Systolic Blood Pressure (mm Hg)	132 (26)	134 (26)	130 (25)
Hypertension (n/%)	2,090 (42.38%)	605 (44.06%)	10,404 (39.3%)
Smoking (n/%)	1,230 (24.94%)	300 (21.85%)	3,893 (24.9%)
Prevalent Heart Failure (n/%)	0 (0%)	0 (0%)	na.
Incident Heart Failure (n/%)	289 (5.86%)	84 (6.118%)	na.
NT-proBNP (pg/ml)	39.26 (56.49)	46.75 (57.2)	na.
MR-proANP (pmol/L)	41.3 (25.2)	43.4 (25.4)	na.
BNP (pg/ml)	12.9 (17.6)	14.9 (19.6)	na.

Quantitative variables are shown as median (interquartile range), quantitative variables as count (proportion).

* Study participants from the FINRISK 1997 cohort

† Study partipants from the FINRISK 1992, FINRISK 2002, FINRISK 2007, NFBC66, HBCS, and Health2000 cohorts

Table 11. Natriuretic peptide variance explained by the genome-wide genotype data.

Trait	Variance Explained	Standard Error
MR-proANP	0.139	0.071
BNP	0.135	0.07
NT-proBNP	0.23	0.072
BNP:NT-proBNP	0.179	0.071

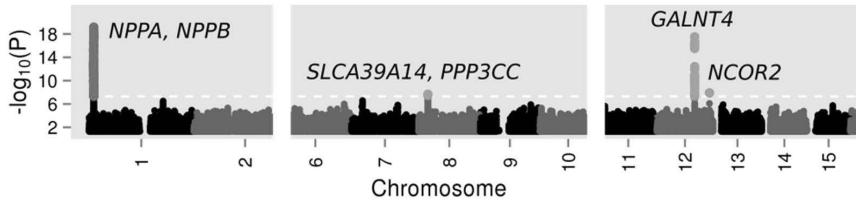

Figure 12. P-values in the genome-wide significant loci for natriuretic peptides. Note: Y-axis cut at Y=18 and data shown for genome-wide significant regions only.

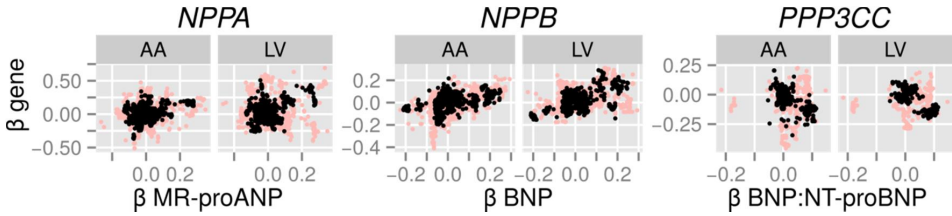

Figure 13. Effects of variants on natriuretic peptides and gene expression. The effect size estimates for the natriuretic peptide traits (X axis) and gene expression (Y axis) are shown for the genes *NPPA*, *NPPB*, and *PPP3CC*. For each gene, the panels titled AA and LV show the values for samples from atrial appendage and left ventricular tissue, respectively. Black dots show the SNPs included in the analysis. Red dots depict SNPs excluded from the analysis based on differences in LD structure between the natriuretic peptide study sample (from Finland) and the gene expression study sample (from USA).

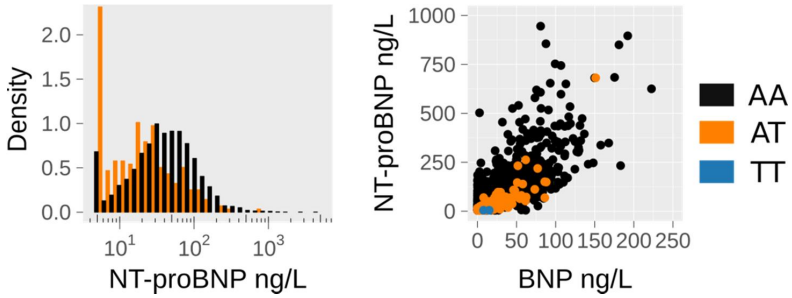

Figure 14. BNP and NT-proBNP distribution by rs61761991 genotype.

Table 12. Genome-wide significant loci for natriuretic peptides

Trait	SNP	Chr	Position	Alleles* (MAF)	Genes (distance, location)	P_{GWAS}	$P_{REPLICATION}$	Beta†	$P_{COMBINED}$†
BNP	rs198379	1	11915467	t/C (0.365)	*NPPB* (3.5kb,3')	6.85×10^{-41}	7.99×10^{-13}	0.249	4.42×10^{-52}
BNP:NT-proBNP	rs61761991	1	11918444	c/T (0.029)	*NPPB* (0.5kb,coding exon)	7.17×10^{-79}	5.71×10^{-26}	1.114	4.81×10^{-103}
	rs7000551	8	22276251	a/G (0.369)	*PPP3CC* (22.5kb,5')	2.16×10^{-8}	0.0248	0.109	2.00×10^{-9}
	rs11105298	12	89876143	t/C (0.211)	*GALNT4* (43.2kb,3')	3.06×10^{-18}	4.11×10^{-6}	0.21	6.77×10^{-23}
MR-proANP	rs3753584	1	11864586	t/C (0.149)	*NPPA* (43.5kb,3')	4.63×10^{-38}	3.48×10^{-7}	0.275	4.19×10^{-13}
	rs701041	12	124999344	G/c (0.106)	*NCOR2* (126kb,intronic)	1.23×10^{-8}	0.9381	-0.088	0.2624
NT-proBNP	rs61761991	1	11918444	c/T (0.029)	*NPPB* (0.5kb,coding exon)	1.72×10^{-51}	5.33×10^{-18}	-0.766	8.76×10^{-68}
	rs10858906	12	89934474	c/T (0.21)	*GALNT4* (15.2kb,5')	1.08×10^{-12}	0.0388	-0.12	3.91×10^{-4}

Association was tested adjusted for geographical sampling region, age, age^2, sex, current smoking status (yes or no), systolic blood pressure, estimated glomerular filtration rate, and genotyping batch.

BNP: B-type natriuretic peptide, MAF: minor allele frequency, MR-proANP: mid-regional proatrial natriuretic peptide, NT-proBNP: aminoterminal pro-B-type natriuretic peptide

* Alleles are given as [reference allele]/[effect allele] with the minor alleles in lower case letters

† Results from the GWAS and replication samples combined using inverse-variance weighted random effects meta-analysis

4.4 STUDY IV: TAKOTSUBO CARDIOMYOPATHY

Takotsubo cadiomyopathy (TTC) is a distinctive condition often presenting with transient heart failure. A genetic predisposition to TTC has been suggested in the literature, but previous genetic studies of TTC have been inconclusive or unconvincing [223-228]. We conducted a GWAS of TTC to investigate the possible contribution of genetic variation to TTC susceptibility and to systematically replicate previously reported results.

Our GWAS did not detect genome-wide significant loci (Table 13). The most statistically significant variant was rs7897505 near the cGMP-dependent protein kinase 1 gene *PRKG1* (rs7897505, odds ratio 0.30, P = 1.04 × 10-6). PRKG1 mediates cardiovascular cGMP signalling and affects cardiomyocyte function, suggesting a plausible biological context for the result [229]. The association is, however, far from statistically significant.

None of the previously reported results replicated in our material. Notably, this included the only previously published genetic association result which has been independently replicated (albeit in a small sample of only 20 TTC patients) [226,227]. The original result was reported by Spinelli et al., who associated the L41Q polymorphism rs17098707 in the G protein-coupled receptor kinase 5 gene *GRK5* with TTC risk [227]. Both the original study and the later replication were performed in Italy, although neither specified the ethnic background of the study participants. However, the allele associated by these studies with TTC is rare in most non-Africans (frequency from 1 to 2 %), but common in at least some African populations (frequency from 24 to 36 %). In a sample of 107 Italians from Tuscany, the frequency of the allele was only 2.3 %. [6] Spinelli et al. reported a surprisingly high frequency of the L41 allele in TTC patients as well as their control sample (376 unaffected volunteers; frequency of the L41 allele 23.9 %), roughly 10-fold the estimate from the Tuscan sample. The results from the two Italian study samples are thus difficult to interpret. The L41 allele was rare and did not associate with risk for TTC in our study, consistent with the results from a study of 92 Australian TTC patients [224].

Prior to our study, one GWAS of TTC has been published by Eitel et al., reporting 18 loci as suggestively (i.e. not genome-wide significantly) associated with TTC [223]. None of these replicated in our study material when the P-values were adjusted for the 18 tests made. When we combined the results from Eitel et al. with those from our study sample using meta-analysis, all but one of the 18 loci had pointwise P-values greater than 0.05 (Table 14). The only variant which was nominally (but not genome-wide) significant in the meta-analysis was rs113154180 (odds ratio=1.97 [95% confidence interval 1.24 to 3.132], P=0.0041). This SNP is located in an intron of the glutamate metabotropic receptor 7 gene *GRM7*.

Given that the onset of TTC is often linked to emotional or psychological stress, the location rs113154180 in *GRM7* is intriguing, as *GRM7* knockout mice have a deficit in an amygdala-dependent fear response. Similarly, blocking GRM7 using a pharmacological antagonist in mice inhibits stress and anxiety-related behavior while reducing amygdala plasticity. Although the association of rs113154180 with TTC is not statistically significant, it is currently the only genetic variant showing some association in two independent studies with genome-wide coverage of genetic variation. Rs113154180 may thus be the top candidate TTC susceptibility locus so far.

Aided by Finnish whole-genome sequencing data from the 1,000 Genomes Project, we estimated the statistical power in our study sample to detect an association for disease alleles with different effect sizes, allele frequencies, and modes of inheritance (Figure 15). This analysis suggested the dataset was well-powered to detect common risk alleles conferring a high relative risk. Two examples of such alleles for other diseases are the APOE epsilon alleles in late-onset Alzheimer's disease (rs4420638; MAF=0.27; OR 95% CI = 3.56 to 4.14) and a regulatory variant near the *RET* gene in Hirschsprung's disease (rs2435357; MAF=0.26; OR 95% CI = 3.2 to 4.7) [230,231]. Statistical power to detect a genome-wide significant association was practically nonexistent for variants with relative risks smaller than 3, regardless of their frequency or mode of inheritance. Power to detect an association for the loci reported by Eitel et al., with less stringent corrections for multiple testing, was adequate (Figure 16). The negative results from the

GWAS, together with the analysis of statistical power, mostly rule out the presence of common, high-risk alleles for TTC in our study population.

Table 13. The most statistically significant results in loci with $P < 5 \times 10^{-5}$ in the GWAS for TTC.

SNP	Chr	Position	Alleles	EAF Cases/Controls	OR (95% CI)	P	Closest Gene
Additive genetic model							
rs199720900	1	203972000	GAA / G	38.78 / 56.47	0.40 (0.26-0.61)	1.74×10^{-5}	NR_027902
rs13022989	2	130792213	G / T	81.63 / 69.43	2.65 (1.67-4.19)	3.29×10^{-5}	NR_026758
rs35388926	3	183611051	T / C	91.13 / 76.12	3.42 (1.96-5.95)	1.44×10^{-5}	*PARL*
rs2069689	5	75912401	G / A	54.44 / 38.81	2.31 (1.55-3.44)	4.21×10^{-5}	*IQGAP2*
rs9350837	6	80668744	G / C	19.36 / 38.27	0.35 (0.22-0.55)	8.75×10^{-6}	*ELOVL4*
rs4631248	6	165390206	T / C	14.45 / 31.35	0.35 (0.22-0.56)	1.74×10^{-5}	*C6orf118*
rs7898868	10	13877152	A / G	65.58 / 47.95	2.66 (1.68-4.20)	3.00×10^{-5}	*FRMD4A*
rs7897505	10	53238703	A / T	20.14 / 38.8	0.30 (0.19-0.49)	1.04×10^{-6}	*PRKG1*
rs1064266	10	116193687	G / C	47.27 / 63.6	0.36 (0.22-0.57)	1.79×10^{-5}	*ABLIM1*
rs11030188	11	3916905	T / C	92.54 / 81.69	4.31 (2.17-8.58)	3.13×10^{-5}	*STIM1*
rs192482099	13	112953585	G / C	86.06 / 74.94	3.34 (1.89-5.89)	3.06×10^{-5}	*LINC01044*
rs8057146	16	85397790	T / C	29.42 / 49.36	0.34 (0.21-0.55)	7.75×10^{-6}	*MIR5093*
rs35841191	17	46890704	GC / G	44.7 / 26.15	2.98 (1.87-4.74)	4.40×10^{-6}	*TTLL6*
rs79756316	18	2304740	A / G	93.17 / 83.86	3.77 (2.05-6.91)	1.82×10^{-5}	*METTL4*
rs8090703	18	6985862	T / A	90.85 / 80.6	3.32 (1.87-5.90)	4.25×10^{-5}	*LAMA1*
Dominant genetic model							
rs1317129	3	176858556	A / G	45.47 / 58.32	0.25 (0.13-0.47)	2.04×10^{-5}	*TBL1XR1*
rs10435058	7	14426831	A / G	38.09 / 47.82	0.14 (0.06-0.35)	1.35×10^{-5}	*DGKB*

Recessive genetic model							
rs571926	1	216162918	A / G	63.06 / 52.97	4.34 (2.17-8.69)	3.27×10^{-5}	*USH2A*
rs2729032	3	2092342	C / G	45.47 / 31.51	3.11 (1.81-5.37)	4.27×10^{-5}	*CNTN4*
rs1886073	6	131489100	A / G	46.14 / 30.6	3.06 (1.79-5.24)	4.59×10^{-5}	*AKAP7*
rs11038791	11	46179754	C / T	56.83 / 44.3	3.95 (2.09-7.46)	2.26×10^{-5}	NR_120565
rs627497	11	64639163	A / G	62.16 / 49.25	4.08 (2.10-7.91)	3.24×10^{-5}	*EHD1*
rs7998695	13	40543697	C / T	56.42 / 40.77	4.11 (2.18-7.77)	1.33×10^{-5}	*COG6*
rs2934477	16	84923247	G / T	50.19 / 37.31	3.40 (1.90-6.10)	3.91×10^{-5}	*CRISPLD2*

Conditional logistic regression matched for gender, genotyping array, and the first two genomic principal components. Chr, chromosome; EAF, effect allele frequency; OR, odds ratio.

* Effect allele/reference allele

† Effect allele frequency in cases/controls

Table 14. Meta-analysis of the loci reported by Eitel et al.

Chr	Pos	SNP	Alleles*	MAF	OR_{FIN}	OR_{EITEL}	P_{FIN}	P_{EITEL}	OR_{META}	P_{META}	I^2†
3	**7605390**	**rs113154180**	**C/T**	**0.158**	**1.545**	**2.48**	**0.0606**	**3.14×10^{-5}**	**1.97 (1.24 to 3.132)**	**0.0041**	**54.696**
3	27815285	rs62253104	C/T	0.35	0.766	0.39	0.1817	1.53×10^{-5}	0.55 (0.284 to 1.065)	0.076	80.829
20	59621304	rs4812257	A/G	0.255	1.162	2.13	0.4635	2.41×10^{-5}	1.586 (0.876 to 2.872)	0.1274	79.786
7	84814246	rs6944978	C/A	0.123	1.069	2.22	0.8162	2.16×10^{-5}	1.592 (0.78 to 3.248)	0.2013	77.842
2	137105542	rs12612435	G/T	0.296	1.096	3.13	0.6597	5.24×10^{-7}	1.844 (0.66 to 5.147)	0.2429	91.343
5	6131250	rs13179382	G/A	0.145	1.027	2.19	0.9181	3.45×10^{-5}	1.529 (0.729 to 3.21)	0.2615	82.326
2	108603502	rs4676168	C/T	0.371	0.986	0.45	0.9417	1.76×10^{-5}	0.666 (0.309 to 1.437)	0.2999	88.616
10	63551773	rs7070797	G/A	0.084	0.939	2.74	0.8565	2.24×10^{-6}	1.662 (0.583 to 4.737)	0.3417	85.421
21	28313238	rs162487	A/T	0.46	1.057	0.493	0.7508	2.42×10^{-5}	0.72 (0.341 to 1.522)	0.3903	90.007
8	87721437	rs4961212	T/G	0.307	0.916	2.07	0.6826	1.99×10^{-5}	1.391 (0.626 to 3.091)	0.4175	88.75
3	112485222	rs1154275	A/G	0.455	0.899	2.09	0.5509	9.02×10^{-6}	1.374 (0.601 to 3.14)	0.4508	91.669
6	122595762	rs72970558	G/A	0.112	0.836	3.07	0.5556	1.30×10^{-5}	1.618 (0.452 to 5.789)	0.4592	90.625
13	32305368	rs4605019	C/T	0.219	0.844	2.86	0.4632	2.97×10^{-5}	1.547 (0.468 to 5.108)	0.4741	92.17
16	5470156	rs12444925	T/A	0.368	1.158	0.505	0.4866	3.88×10^{-5}	0.757 (0.336 to 1.707)	0.5023	89.521
7	22113119	rs17146144	C/T	0.228	0.767	2.27	0.2225	4.29×10^{-5}	1.324 (0.457 to 3.833)	0.6048	92.58
15	59010362	rs56403110	A/T	0.483	1.329	0.516	0.1417	4.66×10^{-5}	0.823 (0.325 to 2.081)	0.6803	92.875
6	6405631	rs9392780	T/G	0.127	0.606	3.19	0.0998	4.45×10^{-6}	1.402 (0.275 to 7.143)	0.6841	94.322
8	22191503	rs13273616	G/C	0.372	0.618	2.22	0.0114	5.94×10^{-6}	1.174 (0.335 to 4.115)	0.8018	95.895

Meta-analysis results of the top 18 loci reported by Eitel et al. from their GWAS of TTC (columns with the subscript EITEL) and the results from the Finnish samples (subscript FIN). The results were combined using a random-effects meta-analysis model.

* Reference allele / effect allele

† I-squared heterogeneity index

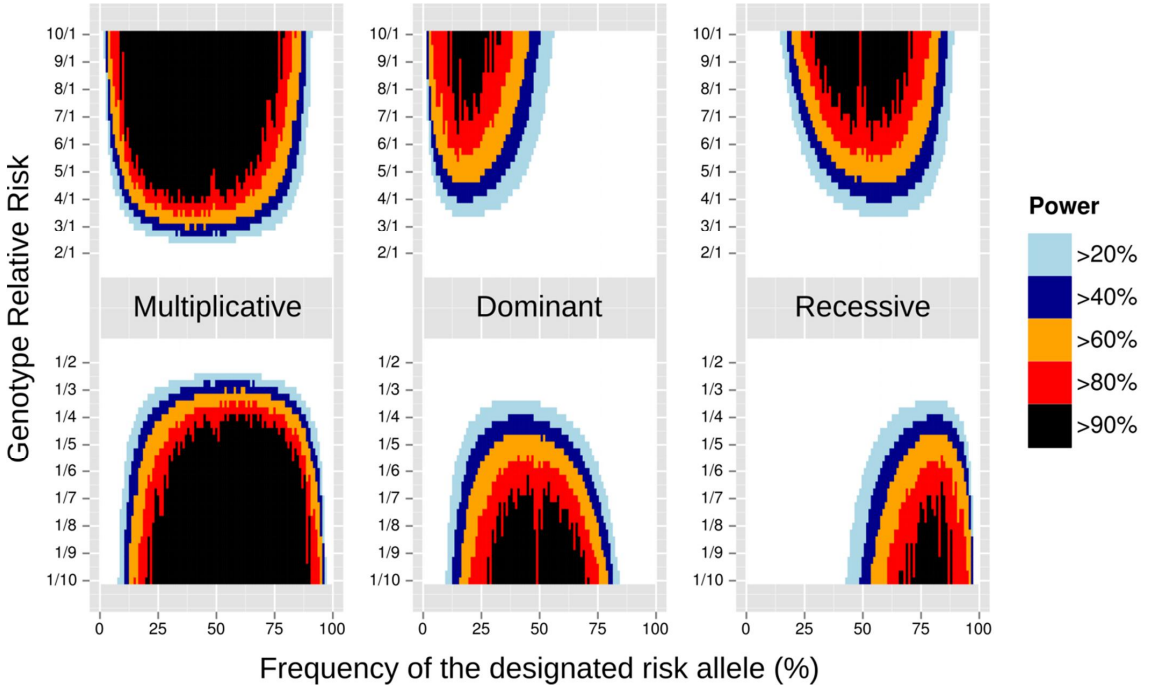

Figure 15. Estimated power to detect a causal allele with genome-wide statistical significance (alpha = 5×10^{-8}). Results are shown separately for the multiplicative, dominant, and recessive modes of inheritance as analyzed using the corresponding correct analysis model. Regions in white show the parameter space with estimated power less than 20%.

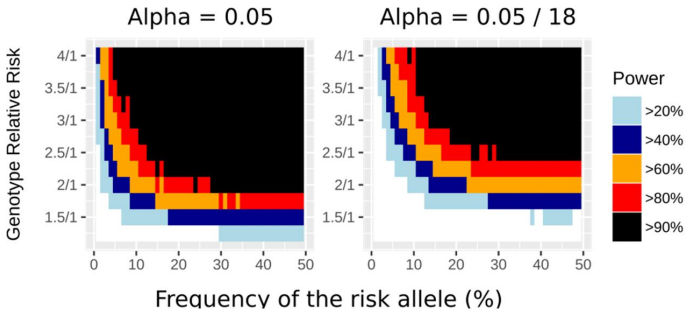

Figure 16. Statistical power in the data for replicating the results of Eitel et al. (alpha = 0.05/18) and for detecting a nominally significant association (alpha = 0.05). Figures are shown assuming a multiplicative mode of inheritance.

5 GENERAL DISCUSSION

5.1 BLOOD PRESSURE AND *PRDM6*

In study I, the working hypothesis predicted shared genetic risk for high blood pressure and intracranial aneurysms. The identification of blood pressure loci had, at that point in time, proven to be surprisingly difficult, even with large sample sizes [125-127]. We suspected that prioritizing variants based on results from a GWAS for IA could be used to detect new blood pressure loci and, at the same time, establish a genetic connection between the two traits.

We successfully detected a novel locus associated with blood pressure near the gene *PRDM6*, at least partly validating our hypothesis. However, this was the only locus of those selected for the study showing such association. Notably, loci which were genome-wide significantly associated with IA in the prior GWAS did not associate with blood pressure. Shared genetic risk for the two traits was thus the exception rather than the rule.

Although the inherent limitations of genetic association studies make it difficult to attribute associations to specific genes, previously published results point to *PRDM6* as a plausible candidate. Recent data from GTEx also show that the blood-pressure-associated SNP is indeed associated with *PRDM6* expression in arterial samples [189]. The role of *PRDM6* in vascular smooth muscle cells suggest that the blood pressure association may stem from an effect on the structure of the arterial wall. Contrary to our speculation in the original publication, the SNP allele associated with elevated blood pressure correlates with decreased *PRDM6* expression in arterial samples. The genetic variant associated with elevated blood pressure may thus be related to decreased vascular smooth muscle cell proliferation.

In the original publication for study I, we suggested that the approach of using analysis results for a related trait (e.g. IA) as a starting point could be a general strategy to identify genetic variants associated with a trait of interest (e.g. blood pressure). This has, however, not been a common trend since the study was published. Rather than modifying study designs, complex trait

genetics has mostly progressed by collecting larger study samples in international consortia. For blood pressure, recent meta-analyses from such studies have covered nearly half a million study participants [232,233].

5.2 STEMI AND NSTEMI

To the author's knowledge, study II is still the only published genome-wide study with stratification of ACS into STEMI and NSTEMI. In fact, remarkably few studies, genetic or otherwise, have tried to identify why some get one of the subtypes rather than the other. This is likely, at least partly, explained by both having the same underlying main cause, coronary artery disease. Anecdotal evidence, based on personal discussions with medical doctors, suggests that many regard the specific type of MI to be caused by somewhat random factors or coincidences, such as the details of plaque rupture. Suggesting that the difference is partly a deterministic result of systematic underlying processes is generally taken as somewhat provocative.

Study II identified a genetic variant which had a much stronger risk for NSTEMI than for STEMI on chromosome 1 near *DRAM2*. The difference was large enough for the signal to meet genome-wide significance only when we analysed the NSTEMI cases. The combined analysis setting, comparing all acute MI cases to the controls, did not detect the association. In a formal comparison, the data was even compatible with the risk being specific to NSTEMI.

The association on chromosome 1 with NSTEMI was not present in large international meta-analyses of coronary artery disease [144,145]. Although they analysed a more heterogeneous phenotype, a reasonable proportion of their study samples had been diagnosed with MI. It seems unlikely these large meta-analyses missed the association solely because they analyzed cases diagnosed with CAD rather than NSTEMI. The association detected in Finns may depend on a genetic variant which is somewhat more common in the Finnish population and too rare elsewhere to be detected using a multinational study sample. Alternatively, the large multinational studies

may have missed the association due to some statistical properties of their much more heterogeneous data.

RNA expression data from whole-blood samples indicated that the NSTEMI association might be related to an effect on the expression of *DRAM2*. Whole blood, however, is probably not the ideal tissue to study this connection and coronary artery or myocardial tissue samples would have been more relevant. For practical reasons, however, they were not available.

A key difference in the pathology of STEMI and NSTEMI is the extent of coronary artery occlusion. In STEMI, it is presumed to be complete, whereas in NSTEMI it is thought to be partial or transient [215,216]. In partial occlusion, a gradient of oxygen availability forms along the area depending on blood flowing through the partially blocked artery. Cells at the top of the gradient near the blocked site suffer less hypoxia, while cells at the bottom receive too little oxygen to survive, leading to the infarction. An intriguing idea is that the genetic NSTEMI association is somehow related to myocardiocytes' capacity to survive during hypoxia. Silencing *DRAM2* has been shown to interfere with cell death *in vitro* in tumor cells, but the relevance of this observation is unclear in the context of differentiated, non-proliferating cardiomyocytes [234].

5.3 NATRIURETIC PEPTIDES AND BLOOD PRESSURE

Natriuretic peptides ANP and BNP were discovered in the 1980's. This is much later than most other hormones - testosterone and estrogen, for example, were discovered half a century before ANP and BNP [235]. Being relative newcomers, the understanding of the physiological functions of ANP and BNP is less complete as compared to more well-known hormones.

Several curious observations have been made for ANP and BNP. In mice, overexpression of either *NPPA* or *NPPB* triggers hypotension, and knocking out *NPPA* leads to hypertension [236-238]. Surprisingly, knocking out *NPPB* triggers cardiac fibrosis, not hypertension [239]. In humans, the effects of

ANP or BNP infusions vary based on baseline status. In patients suffering from HF, infusions of ANP or BNP trigger various changes, including a decrease in arterial pressure, but in healthy males only induce natriuresis with no effect on arterial BP [219,240-242]. In those suffering from essential hypertension, both ANP and BNP have been reported to be hypotensive with BNP surprisingly showing up to threefold greater potency than ANP, even though both have similar receptor affinities [243,244]. In obese men, ANP and blood pressure show no association or even an unexpected negative association [245].

Our results show that genetic differences in circulating ANP concentrations had a small but significant effect on arterial blood pressure. The combined effect of the three MR-proANP associated SNPs identified in the study associated with a significant difference in risk for hypertension. Prior studies had shown a link between genetic variation near *NPPA* and *NPPB*, but due to the limitations of their data, they have not been able to attribute the associations to either [109]. Our data, combined with gene expression data from myocardial samples, show that genetic variants increasing the expression of *NPPA* in the myocardium associate with greater circulating concentration of MR-proANP, lower arterial blood pressure, and reduced risk of hypertension in the general population.

Contrary to ANP, our results did not show an association with genetic differences in circulating BNP concentration and arterial blood pressure. This, however, does not mean that BNP has no such effect. Being a study of naturally occurring genetic variation, the results are limited to observing whatever genetic variants were present in our study population. The effect of the BNP-associated SNP may have been too small to cause a large enough difference in blood pressure to be detected in our study. That said, data on the effects of BNP in the general population are scarce. Previously published genetic studies near *NPPA* and *NPPB* have been used in the literature ambiguously to make a case for both ANP and BNP as blood pressure regulators [246,247]. Our study shows that the genetic results only specifically support ANP and are, at most, uninformative for BNP.

In addition to the *NPPA-NPPB* locus, three loci elsewhere in the genome have been associated with circulating natriuretic peptide concentrations, including the *PPP3CC* locus discovered in study III [107]. Our results show that a common characteristic of these loci is their association to NT-proBNP rather than C-terminal BNP. For the *PPP3CC* locus, the association was strongest for the ratio of BNP to NT-proBNP. These loci may be related to the processing of proBNP in the circulation. Contrary to proANP, which is cleaved by the cardiomyocytes secreting the hormone, proBNP is cleaved peripherally and a large proportion of circulating BNP is actually contained in uncleaved proBNP [101,248]. Nonspecific mechanisms related to the elimination of circulating polypeptides from the blood, e.g. kidney function, might influence BNP:NT-proBNP ratio. However, the loci did not associate with MR-proANP. The unknown mechanism(s) related to the pronounced association with NT-proBNP rather than BNP are specific enough not to affect circulating MR-proANP.

Rs61761991 in N-terminal proBNP associated with a dramatic decrease in measured NT-proBNP concentration, with no effect on BNP. As the SNP is located in the region used as the antigen to develop antibodies for NT-proBNP immunoassays, the simplest explanation is that NT-proBNP antibodies have low affinity to the rarer rs61761991 allele [186]. While the SNP is very rare elsewhere, its minor allele frequency is approximately 2% to 3% in Finland. Roughly one out of twenty Finns are thus heterozygous for rs61761991 and may present with artefactually low NT-proBNP concentrations, possibly causing mistaken rule-out of suspected heart failure. The unluckiest one out of thousand Finnish individuals are rare rs61761991 homozygotes and will present essentially zero values for NT-proBNP assays. The potential effect on diagnosis of HF would be eliminated by using C-terminal BNP rather than NT-proBNP assays in the Finnish population.

5.4 TAKOTSUBO CARDIOMYOPATHY

A GWAS of Takotsubo cardiomyopathy in study IV did not identify genome-wide significant loci and showed a nonreplication for loci reported by prior

studies. The study sample was limited in size, and we estimated that we had adequate power to detect only common, high-risk disease alleles in the genome-wide setting. The situation was better for the targeted replication of previously reported results, where multiple testing burden was less severe.

Scientific progress on Takotsubo cardiomyopathy, the curious form of transient heart failure, has been severely hampered by small patient cohorts. This is the result of TTC being a rare, late-onset disease, which has only quite recently been described. In fact, a large part of the modern literature on TTC consists of case studies describing single patients rather than cohorts of any size. Compared to the status quo, the cohort investigated in study IV is an improvement. The study sample was, however, far from large enough to identify disease alleles with low penetrances and effect sizes typical of complex diseases, where odds ratios are with few exceptions below 1.5.

Given that data on TTC is extremely limited, scientists should extract as much utility from the data at hand as possible. In the meta-analysis of the recently published GWAS of TTC, and the results from study IV, one locus was nominally statistically significant [223]. Although this SNP in the glutamate metabotropic receptor 7 gene *GRM7* is not statistically significant after accounting for the number of variants in the meta-analysis, it is the top result in the combined analysis of the two most extensive genetic studies of TTC to date. As *GRM7* is related to fear response and stressful situations are known to be among the triggers of TTC, *GRM7* may be the best candidate for molecular studies in TTC, regardless of the SNP association failing to meet strict statistical significance [249,250].

5.5 LIMITATIONS AND FUTURE PROSPECTS

The study of human disease is ultimately aimed at producing effective preventive strategies or treatments. In this respect, the utility of the results from studies I to III is limited by the inability to conclusively attribute the observed trait associations to specific genetic variants and genes. Despite our attempts at deciphering these associations, success was limited and leaves

much to be desired. The findings do not unambiguously point to specific biological mechanisms which could be targeted in hypertension or myocardial infarction to, ultimately, ease the population burden of heart failure. This limitation is shared by most results from GWA studies, with rare exceptions, such as the NT-proBNP SNP rs61761991, where the interpretation of the results is easier mostly by coincidence. Addressing this limitation systematically to reap the full benefits from GWA studies will probably require the development of entirely novel analytical or experimental approaches.

The GWA analysis of Takotsubo cardiomyopathy in study IV was limited by its small sample size. The study sample was not statistically well powered to identify genome-wide significant associations to genetic variants with modest or small effect sizes. Future studies of TTC will require substantially larger samples to address genetic variation typical of complex disease.

The results of GWA studies are often quantified in terms of the fraction of heritability explained by all known loci or the newly identified loci alone. These estimates can then be used to assess whether new studies of the same traits are justified. The samples investigated in this thesis are not large enough to produce such estimates with meaningful precision for the loci identified in the present studies. However, the fraction can be expected to be small, as the loci are few and their effect sizes are modest. In the future, much larger GWA studies will almost certainly identify new loci. Although ever larger GWA studies have been criticized, and valid questions of their potential value have been raised, they represent a rare simple path forward among the complexities of biology [251]. Even if their value cannot be guaranteed, future GWA studies of millions of people will be a reasonable use of scientific resources.

6 CONCLUSIONS

In study I, we identified rs2287696 in *PRDM6* as associated with blood pressure and risk for intracranial aneurysms. Based on prior reports on the biological function of *PRDM6*, we speculated, in the original study, that the association was related to the level of *PRDM6* expression in the arterial wall. Recently published data supports a link between rs2287696 and arterial *PRDM6* expression.

Study II, a GWAS of MI, identified a locus on chromosome 1 near *DRAM2* associated with increased risk for NSTEMI with a much weaker effect for STEMI. This result was an outlier among variants previously associated with risk for coronary artery disease, which did not show marked differences between NSTEMI and STEMI in our study sample. The NSTEMI association is likely related to pathomechanisms other than the onset of coronary artery disease.

The genome-wide analysis in study III associated several variants in three loci near *NPPA-NPPB*, *GALNT4,* and *PPP3CC* with the circulating concentration of natriuretic peptides. Variants associated with increased circulating MR-proANP also correlated with lower blood pressure and lower risk of hypertension. Variants associated with BNP or NT-proBNP did not associate with blood pressure. The genetic data thus supports the role of ANP in blood pressure regulation in the general population but does not offer similar support for BNP.

Rs61761991 in N-terminal proBNP associated with a dramatic decrease in measured NT-proBNP concentration, with no effect on BNP. The association is likely caused by a reduced affinity of the NT-proBNP assay antibody to the rarer rs61761991 allele. Roughly one out of twenty Finns are heterozygous for rs61761991, possibly causing mistaken rule out of suspected heart failure. This risk would be eliminated by using C-terminal BNP rather than NT-proBNP assays in the Finnish population for diagnosis of HF.

The GWA study of Takotsubo cardiomyopathy did not identify susceptibility loci. Furthermore, previously reported associations did not

replicate in our study sample. Taken together with the estimates of statistical power for our study sample, these results indicate that there are no common alleles conferring moderate to high risk for TTC in the Finnish population. The results also mostly refute the associations published in previous studies. If genetic predisposition to TTC does exist, it is likely due to variants with low penetrance, low allele frequency, or both.

7 ACKNOWLEDGEMENTS

The studies forming this thesis were carried out at the National Institute for Health and Welfare THL, Helsinki, Finland and at the Institute for Molecular Medicine Finland FIMM, University of Helsinki. I wish to thank Juhani Eskola and Jaakko Kaprio, the directors of THL and FIMM, and Anu Jalanko, head of the Genomics and Biomarkers Unit at THL, for providing superb research facilities. All participants of the cohorts analyzed are thanked for making the studies possible.

I am grateful for the fincancial support of the Integrative Life Science Doctoral Programme of the University of Helsinki, Aarne Koskelo Foundation, Biomedicum Helsinki Foundation, the Finnish Foundation for Cardiovascular Research, Ida Montin Foundation, Paulo Foundation, and Finnish Cultural Fund during my academic career. The importance of this support cannot be overstated.

Research Professor Markus Perola supervised my thesis with great patience and expertise. Working with him has been a privilege and a pleasure: with his experience and skill he has guided my work while giving me the necessary room for independent scientific exploration and growth. He has succeeded in building a great team to work in - being a part of that team has made my time as a doctoral candidate unforgettable. Thank you, Markus!

Associate Professor Mikko Pietilä and Professor Arto Mannermaa pre-examined my thesis. I thank both for taking the time to critically and constructively review the work. Your expert remarks and suggestions were priceless in improving the manuscript. After the rather lonesome task of writing the *opus*, receiving your insightful comments also reminded me that I did not write it for myself but as a part of a scientific tradition. I wish to thank Professor Juha Sinisalo for accepting the role of custos of my thesis defence and Professor Katriina Aalto-Setälä for accepting the role of the opponent. My thesis committee members, Professor Hannes Lohi and Associate Professor Iiris Hovatta are also warmly thanked: our meetings were equally few as they were important.

I am indebted to my great collaborators and coauthors. Veikko Salomaa is thanked for his vast epidemiological knowledge and helpful comments on the manuscripts. Aki Havulinna was indispensable in matters related to the population datasets and their statistical modeling. Taru Tukiainen lended her brilliant scientific skills and enthusiasm to the natriuretic peptide study just when it was most needed. Emília Gaál-Paavola set the example of a determined and dedicated researcher as she took me in as a novice scientist to the blood pressure study. Satu Vaara is thanked for engaging discussions regarding (N)STEMI. Marja-Liisa Lokki provided useful and timely comments on the manuscripts. Olavi Parkkonen has been a perfect teammate in matters related to the traditional octopus trap disease of the heart (TOTDH). Johannes Kettunen helped navigate many data-related issues. Antti-Pekka Sarin imputed mostly every piece of data available. Matti Pirinen helped to answer some very important statistical questions. Kati Kristiansson was indispensable both as a coauthor as a mentor. Aarno Palotie and Samuli Ripatti have promoted excellence in human genetics throughout FIMM. Tiina Paunio, Olli Raitakari, Terho Lehtimäki, Johan Eriksson, Pekka Karhunen, Kjell Nikus, Markku Eskola, Markku S. Nieminen and Juha Sinisalo have successfully steered indispensable population cohorts and provided expert comments on the manuscripts. Collaboration with Mikko Pakarinen at the Children's hospital has been highly motivating and an important part of my scientific education even though the projects were not directly related to my thesis studies. Likewise, not direcly related to the PhD thesis work but a very rewarding part of these years nonetheless, Marja Puurunen and Kaija Javela at the Finnish Red Cross Blood Service are thanked for their collaboration in the antithrombin genetics project.

Group Perola at THL was my home away from home during the past few years (how many are they now?). Anni has been there with me from the very beginning. Smart, talented, dependable - she has been the perfect colleague and a friend. Anni: this time would not have been the same without you sharing the ups, downs, and slopes of science and life! Tero, a scholar, a gentleman, and a true friend, has given me some of the best pieces of advice of my entire life and taught me to aim at *superiority* in everything. Iiro, a

unique source of enthusiasm, odd music choices, and great conversations, has been an overall awesome person. Marjis is thanked for her wit, personality, and company in the office and at the club, until suddenly disappearing into the dark corners of Häme. Mikko has a great weird sense of humor, it has been a pleasure to hang around with him. Joe answered every single question in statistical genetics I asked him and managed to teach me important lessons while doing so. Together with Hannele, Liisa, Heikki, Nita, Jasmine, Emma, Kirsi, Annika, Outi, Peter, Minna, Valtter, Birgit, and Marjo they have made the Perola group such an amazing team. Thank you!

I wish to thank all the wonderful people at the Genomics and Biomarkers unit of THL and at THL biobank. Tuija Holm and Liisa Penttilä assisted me with countless of things and were always ready to provide expert help when needed. Anne Vikman is thanked for laboratory mastery. Marketta Taimi patiently and carefully helped to make many things the right way. Päivi Laiho, Tuuli Sistonen, Essi Kaiharju and Minttu Sauramo are thanked for knowing what the right samples and projects were even when I did not. Niina, Matti, Katri, Aija, Alexandra, Johanna, Anne, Hanski, Marja-Riitta, Kaisa, Sonja, Arja, Auli, Teppo, Jarkko, P-P, Markus, Tea, and Minna are thanked for their great company. It has been a pleasure to work with you!

I would like to thank all the people at FIMM, my second academic home, for making it the special place it is: Annu, Ansku, Diana, Emmi, Elina, Elisabeth, Hannele, Himanshu, Jarmo, Javier, Jenni, Kalle, Virpi, Mervi, Marjo, Mikko, Jarmo, Paavo, Pekka, Peter, Pietari, Priit, Pyry, Sini, Susanna, Sharif, Timo, Olli, Olle, and Luca. Special thanks to the people I shared the office with: Ida, Jaakko, Christian, Om, Rubina, Vishal, Teemu, Meri, Juha, and Tiia.

Family and friends: thank you for everything.

8 REFERENCES

1. GBD 2016 Causes of Death Collaborators. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: A systematic analysis for the global burden of disease study 2016. *Lancet.* 2017;390:1151-210.

2. Lloyd-Jones DM, Larson MG, Leip EP, Beiser A, D'Agostino RB, Kannel WB, et al. Lifetime risk for developing congestive heart failure: The Framingham heart study. *Circulation.* 2002;106:3068-72.

3. Ziaeian B, Fonarow GC. Epidemiology and aetiology of heart failure. *Nat. Rev. Cardiol.* 2016;13:368-78.

4. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JG, Coats AJ, et al. 2016 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure: The task force for the diagnosis and treatment of acute and chronic heart failure of the European society of cardiology (ESC) developed with the special contribution of the heart failure association (HFA) of the ESC. *Eur. Heart J.* 2016;37:2129-200.

5. Peltonen L, Jalanko A, Varilo T. Molecular genetics the Finnish disease heritage. *Hum. Mol. Genet.* 1999;8:1913-23.

6. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature.* 2015;526:68-74.

7. Bamshad M, Wooding S, Salisbury BA, Stephens JC. Deconstructing the relationship between genetics and race. *Nature Reviews Genetics.* 2004;5:598-609.

8. Avery OT, Macleod CM, McCarty M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types : Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type iii. *J. Exp. Med.* 1944;79:137-58.

9. Watson JD, Crick FHC. A structure for deoxyribose nucleic acid. *Nature.* 1953;737-8.

10. Nirenberg MW, Matthaei JH. The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci. U. S. A.* 1961;47:1588-602.

11. Holley RW, Apgar J, Everett GA, Madison JT, Marquisee M, Merrill SH, et al. Structure of a ribonucleic acid. *Science.* 1965;1462-5.

12. Strachan T, Read A. Human molecular genetics. 4th ed. New York: Garland Science; 2011.

13. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. *Nature.* 2006;444:444-54.

14. Hardy GH. Mendelian proportions in a mixed population. *Science.* 1908;28:49-50.

15. Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. A fine-scale map of recombination rates and hotspots across the human genome. *Science.* 2005;310:321-4.

16. Lewontin R, Kojima K. The evolutionary dynamics of complex polymorphisms. *Evolution.* 1960;14:458-72.

17. Lewontin RC. The interaction of selection and linkage. I. general considerations; heterotic models. *Genetics.* 1964;49:49-67.

18. Hill W, Robertson A. Linkage disequilibrium in finite populations. *TAG Theoretical and Applied Genetics.* 1968;38:226-31.

19. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. 2001;.

20. Church DM, Schneider VA, Graves T, Auger K, Cunningham F, Bouk N, et al. Modernizing reference genome assemblies. *PLoS biology.* 2011;9:e1001091.

21. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: The reference human genome annotation for the ENCODE project. *Genome Res.* 2012;22:1760-74.

22. Babayev E, Seli E. Oocyte mitochondrial function and reproduction. *Curr. Opin. Obstet. Gynecol.* 2015;27:175-81.

23. Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, et al. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature.* 2011;478:476-82.

24. Mills RE, Bennett EA, Iskow RC, Devine SE. Which transposable elements are active in the human genome? *Trends in genetics.* 2007;23:183-91.

25. Gerstein MB, Bruce C, Rozowsky JS, Zheng D, Du J, Korbel JO, et al. What is a gene, post-ENCODE? history and updated definition. *Genome Res.* 2007;17:669-81.

26. Pruitt KD, Brown GR, Hiatt SM, Thibaud-Nissen F, Astashyn A, Ermolaeva O, et al. RefSeq: An update on mammalian reference sequences. *Nucleic Acids Res.* 2013;42:D756-63.

27. NCBI homo sapiens annotation release 108. Available from: https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Homo_sapiens/108/.

28. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012;489:57-74.

29. Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, Cline MS, et al. The UCSC genome browser database: Update 2011. *Nucleic Acids Res.* 2010;39:D876-82.

30. Jaenisch R, Bird A. Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. *Nat. Genet.* 2003;33:245-54.

31. Rothbart SB, Strahl BD. Interpreting the language of histone and DNA modifications. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms.* 2014;1839:627-43.

32. Kim T, Shiekhattar R. Architectural and functional commonalities between enhancers and promoters. *Cell.* 2015;162:948-59.

33. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. The NIH roadmap epigenomics mapping consortium. *Nat. Biotechnol.* 2010;28:1045-8.

34. ENCODE Project Consortium. The ENCODE (ENCyclopedia of DNA elements) project. *Science.* 2004;306:636-40.

35. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-André V, Sigova AA, et al. Super-enhancers in the control of cell identity and disease. *Cell.* 2013;155:934-47.

36. Sherry ST, Ward M, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: The NCBI database of genetic variation. *Nucleic Acids Res.* 2001;29:308-11.

37. 1000 Genomes Project. Variation in genome-wide mutation rates within and between human families. *Nat. Genet.* 2011;43:712-4.

38. Fu W, O'connor TD, Jun G, Kang HM, Abecasis G, Leal SM, et al. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature.* 2013;493:216-20.

39. Schlebusch CM, Malmström H, Günther T, Sjödin P, Coutinho A, Edlund H, et al. Ancient genomes from southern Africa pushes modern human divergence beyond 260,000 years ago. *bioRxiv.* 2017;145409.

40. Hublin J, Ben-Ncer A, Bailey SE, Freidline SE, Neubauer S, Skinner MM, et al. New fossils from Jebel Irhoud, Morocco and the pan-African origin of homo sapiens. *Nature.* 2017;546:289-92.

41. Richter D, Grün R, Joannes-Boyau R, Steele TE, Amani F, Rué M, et al. The age of the hominin fossils from Jebel Irhoud, Morocco, and the origins of the middle stone age. *Nature.* 2017;546:293-6.

42. Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, Cavalli-Sforza LL. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc. Natl. Acad. Sci. U. S. A.* 2005;102:15942-7.

43. Henke W, Henke W, Tattersall I, editors. Handbook of paleoanthropology. 2nd ed. New York: Springer Reference; 2015.

44. Veeramah KR, Hammer MF. The impact of whole-genome sequencing on the reconstruction of human population history. *Nature Reviews Genetics.* 2014;15:149-62.

45. McEvoy BP, Powell JE, Goddard ME, Visscher PM. Human population dispersal "out of Africa" estimated from linkage disequilibrium and allele frequencies of SNPs. *Genome Res.* 2011;21:821-9.

46. Deshpande O, Batzoglou S, Feldman MW, Cavalli-Sforza LL. A serial founder effect model for human settlement out of Africa. *Proc. Biol. Sci.* 2009;276:291-300.

47. Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, et al. Worldwide human relationships inferred from genome-wide patterns of variation. *Science.* 2008;319:1100-4.

48. Sjödin P, E. Sjöstrand A, Jakobsson M, Blum MG. Resequencing data provide no evidence for a human bottleneck in Africa during the penultimate glacial period. *Mol. Biol. Evol.* 2012;29:1851-60.

49. Gignoux CR, Henn BM, Mountain JL. Rapid, global demographic expansions after the origins of agriculture. *Proc. Natl. Acad. Sci. U. S. A.* 2011;108:6044-9.

50. Historical estimates of world population. Available from: https://www.census.gov/population/international/data/worldpop/table_history.php.

51. United Nations. Table 1. In: 2015 Demographic Yearbook. 66th ed. New York: United Nations; 2016. p. 46.

52. Hartl DL, Clark AG. Principles of population genetics. 4th ed. Sunderland: Sinauer Associates; 2007.

53. Keinan A, Clark AG. Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science.* 2012;336:740-3.

54. The International HapMap Consortium. The international HapMap project. *Nature.* 2003;426:789-96.

55. Varilo T. The age of the mutations in the Finnish disease heritage; a genealogical and linkage disequilibrium study [dissertation]. University of Helsinki; 1999.

56. Sajantila A, Salem AH, Savolainen P, Bauer K, Gierig C, Paabo S. Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population. *Proc. Natl. Acad. Sci. U. S. A.* 1996;93:12035-9.

57. Lim ET, Würtz P, Havulinna AS, Palta P, Tukiainen T, Rehnström K, et al. Distribution and medical impact of loss-of-function variants in the Finnish founder population. *PLoS genetics.* 2014;10:e1004494.

58. Kerminen S, Havulinna AS, Hellenthal G, Martin AR, Sarin A, Perola M, et al. Fine-scale genetic structure in Finland. *G3: Genes|Genomes|Genetics.* 2017;7:3459-68.

59. Lappalainen T, Koivumäki S, Salmela E, Huoponen K, Sistonen P, Savontaus M, et al. Regional differences among the Finns: A Y-chromosomal perspective. *Gene.* 2006;376:207-15.

60. Kittles RA, Perola M, Peltonen L, Bergen AW, Aragon RA, Virkkunen M, et al. Dual origins of Finns revealed by Y chromosome haplotype variation. *The American Journal of Human Genetics.* 1998;62:1171-9.

61. Fisher RA. XV.—The correlation between relatives on the supposition of mendelian inheritance. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh.* 1919;52:399-433.

62. Lush JL. Intra-sire correlations or regressions of offspring on dam as a method of estimating heritability of characteristics. *Proceedings of the American Society of Animal Nutrition.* 1940;1940:293-301.

63. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era--concepts and misconceptions. *Nat. Rev. Genet.* 2008;9:255-66.

64. Polderman TJ, Benyamin B, De Leeuw CA, Sullivan PF, Van Bochoven A, Visscher PM, et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* 2015;47:702-9.

65. Dudley JT, Kim Y, Liu L, Markov GJ, Gerold K, Chen R, et al. Human genomic disease variants: A neutral evolutionary explanation. *Genome Res.* 2012;22:1383-94.

66. Cooper DN, Krawczak M, Polychronakos C, Tyler-Smith C, Kehrer-Sawatzki H. Where genotype is not predictive of phenotype: Towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum. Genet.* 2013;132:1077-130.

67. Sham PC. Statistics in human genetics. Chichester: Wiley; 1998.

68. Szumilas M. Explaining odds ratios. *J. Can. Acad. Child. Adolesc. Psychiatry.* 2010;19:227-9.

69. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online mendelian inheritance in man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* 2005;33:D514-7.

70. Blekhman R, Man O, Herrmann L, Boyko AR, Indap A, Kosiol C, et al. Natural selection on genes that underlie human disease susceptibility. *Current biology.* 2008;18:883-9.

71. Cai JJ, Borenstein E, Chen R, Petrov DA. Similarly strong purifying selection acts on human disease genes of all evolutionary ages. *Genome biology and evolution.* 2009;1:131-44.

72. Gibson G. Rare and common variants: Twenty arguments. *Nat. Rev. Genet.* 2012;13:135-45.

73. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science.* 1996;273:1516-7.

74. Psychiatric GWAS Consortium Coordinating Committee. Genomewide association studies: History, rationale, and prospects for psychiatric disorders. *Am. J. Psychiatry.* 2009;5:549-56.

75. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI catalog of published genome-wide association studies (GWAS catalog). *Nucleic Acids Res.* 2017;45:D896-901.

76. Saunders AM, Strittmatter WJ, Schmechel D, George-Hyslop PH, Pericak-Vance MA, Joo SH, et al. Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease. *Neurology.* 1993;43:1467-72.

77. Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers E, Berk R, et al. Redefine statistical significance. *Nature Human Behaviour.* 2018;2:6-10.

78. Wasserstein RL, Lazar NA. The ASA's statement on p-values: Context, process, and purpose. *The American Statistician.* 2016;70:129-33.

79. Hart AB, de Wit H, Palmer AA. Candidate gene studies of a promising intermediate phenotype: Failure to replicate. *Neuropsychopharmacology.* 2013;38:802-16.

80. Munafo M. Candidate gene studies in the 21st century: Meta-analysis, mediation, moderation. *Genes, Brain and Behavior.* 2006;5:3-8.

81. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. *Nat. Genet.* 2001;29:306-9.

82. Morgan TM, Krumholz HM, Lifton RP, Spertus JA. Nonvalidation of reported genetic risk factors for acute coronary syndrome in a large-scale replication study. *JAMA.* 2007;297:1551-61.

83. LaFramboise T. Single nucleotide polymorphism arrays: A decade of biological, computational and technological advances. *Nucleic Acids Res.* 2009;37:4181-93.

84. Terwilliger JD, Göring HH. Gene mapping in the 20th and 21st centuries: Statistical methods, data analysis, and experimental design. *Human biology.* 2009;81:663-728.

85. Hiekkalinna T. On the superior power of likelihood-based linkage disequilibrium mapping in large multiplex families compared to population based case-control designs [dissertation]. University of Helsinki; 2012.

86. Haplotype Reference Consortium. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* 2016;48:1279-83.

87. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet. Epidemiol.* 2008;32:381-5.

88. Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nature Reviews Genetics.* 2010;11:499-511.

89. Browning BL, Browning SR. Genotype imputation with millions of reference samples. *The American Journal of Human Genetics.* 2016;98:116-26.

90. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* 2012;44:955-9.

91. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *The American Journal of Human Genetics.* 2012;90:7-24.

92. Metra M, Teerlink JR. Heart failure. *Lancet.* 2017;390:1981-95.

93. Mosterd A, Hoes AW. Clinical epidemiology of heart failure. *Heart.* 2007;93:1137-46.

94. Bleumink GS, Knetsch AM, Sturkenboom MC, Straus SM, Hofman A, Deckers JW, et al. Quantifying the heart failure epidemic: Prevalence, incidence rate, lifetime risk and prognosis of heart failure: The rotterdam study. *Eur. Heart J.* 2004;25:1614-9.

95. Redfield MM, Jacobsen SJ, Burnett Jr JC, Mahoney DW, Bailey KR, Rodeheffer RJ. Burden of systolic and diastolic ventricular dysfunction in the community: Appreciating the scope of the heart failure epidemic. *JAMA.* 2003;289:194-202.

96. Celermajer DS, Chow CK, Marijon E, Anstey NM, Woo KS. Cardiovascular disease in the developing world. *J. Am. Coll. Cardiol.* 2012;60:1207-16.

97. Dokainish H, Teo K, Zhu J, Roy A, AlHabib KF, ElSayed A, et al. Heart failure in Africa, asia, the middle east and south america: The INTER-CHF study. *Int. J. Cardiol.* 2016;204:133-41.

98. de Bold AJ. Atrial natriuretic factor: A hormone produced by the heart. *Science.* 1985;230:767-70.

99. Sudoh T, Kangawa K, Minamino N, Matsuo H. A new natriuretic peptide in porcine brain. *Nature.* 1988;332:78-81.

100. Yasue H, Yoshimura M, Sumida H, Kikuta K, Kugiyama K, Jougasaki M, et al. Localization and mechanism of secretion of B-type natriuretic peptide in comparison with those of A-type natriuretic peptide in normal subjects and patients with heart failure. *Circulation.* 1994;90:195-203.

101. Kerkela R, Ulvila J, Magga J. Natriuretic peptides in the regulation of cardiovascular physiology and metabolic events. *J. Am. Heart Assoc.* 2015;4:e002423.

102. Dao Q, Krishnaswamy P, Kazanegra R, Harrison A, Amirnovin R, Lenert L, et al. Utility of B-type natriuretic peptide in the diagnosis of congestive heart failure in an urgent-care setting. *J. Am. Coll. Cardiol.* 2001;37:379-85.

103. Burnett Jr J, Kao P, Hu D, Heser D, Heublein D, Granger J, et al. Atrial natriuretic peptide elevation in congestive heart failure in the human. *Science.* 1986;231:1145-8.

104. Maisel A, Mueller C, Adams K, Anker SD, Aspromonte N, Cleland JG, et al. State of the art: Using natriuretic peptide levels in clinical practice. *European journal of heart failure.* 2008;10:824-39.

105. Del Greco MF, Pattaro C, Luchner A, Pichler I, Winkler T, Hicks AA, et al. Genome-wide association analysis and fine mapping of NT-proBNP level provide novel insight into the role of the MTHFR-CLCN6-NPPA-NPPB gene cluster. *Hum. Mol. Genet.* 2011;20:1660-71.

106. Folkersen L, Fauman E, Sabater-Lleal M, Strawbridge RJ, Frånberg M, Sennblad B, et al. Mapping of 79 loci for 83 plasma protein biomarkers in cardiovascular disease. *PLoS Genetics.* 2017;13:e1006706.

107. Johansson A, Eriksson N, Lindholm D, Varenhorst C, James S, Syvanen AC, et al. Genome-wide association and mendelian randomization study of NT-proBNP in patients with acute coronary syndrome. *Hum. Mol. Genet.* 2016;25:1447-56.

108. Musani SK, Fox ER, Kraja A, Bidulescu A, Lieb W, Lin H, et al. Genome-wide association analysis of plasma B-type natriuretic peptide in blacks: The Jackson heart study. *Circ. Cardiovasc. Genet.* 2015;8:122-30.

109. Newton-Cheh C, Larson MG, Vasan RS, Levy D, Bloch KD, Surti A, et al. Association of common variants in NPPA and NPPB with circulating natriuretic peptides and blood pressure. *Nat. Genet.* 2009;41:348-53.

110. Pereira NL, Tosakulwong N, Scott CG, Jenkins GD, Prodduturi N, Chai Y, et al. Circulating atrial natriuretic peptide genetic association study identifies a novel gene cluster associated with reduced NT-proANP, increased stroke and higher diastolic blood pressure. *BMC Pharmacology and Toxicology.* 2015;16:A37.

111. Jabre P, Jouven X, Adnet F, Thabut G, Bielinski SJ, Weston SA, et al. Atrial fibrillation and death after myocardial infarction: A community study. *Circulation.* 2011;123:2094-100.

112. Cowie M, Wood D, Coats A, Thompson S, Poole-Wilson P, Suresh V, et al. Incidence and aetiology of heart failure; a population-based study. *Eur. Heart J.* 1999;20:421-8.

113. Hazebroek M, Dennert R, Heymans S. Idiopathic dilated cardiomyopathy: Possible triggers and treatment strategies. *Netherlands Heart Journal.* 2012;20:332-5.

114. Velagaleti RS, Vasan RS. Heart failure in the twenty-first century: Is it a coronary artery disease or hypertension problem? *Cardiology Clinics.* 2007;25:487-95.

115. Gheorghiade M, Sopko G, De Luca L, Velazquez EJ, Parker JD, Binkley PF, et al. Navigating the crossroads of coronary artery disease and heart failure. *Circulation.* 2006;114:1202-13.

116. McMurray JJ, Stewart S. Epidemiology, aetiology, and prognosis of heart failure. *Heart.* 2000;83:596-602.

117. Wong ND. Epidemiological studies of CHD and the evolution of preventive cardiology. *Nature Reviews Cardiology.* 2014;11:276-89.

118. Council ES, Redon J, Narkiewicz K, Nilsson PM, Burnier M, Viigimaa M, et al. 2013 ESH/ESC guidelines for the management of arterial hypertension. *Eur. Heart J.* 2013;34:2159-219.

119. Fagard R, Brguljan J, Staessen J, Thijs L, Derom C, Thomis M, et al. Heritability of conventional and ambulatory blood pressures. A study in twins. *Hypertension.* 1995;26:919-24.

120. Adrogué HJ, Madias NE. Sodium and potassium in the pathogenesis of hypertension. *N. Engl. J. Med.* 2007;356:1966-78.

121. Intersalt: An international study of electrolyte excretion and blood pressure. results for 24 hour urinary sodium and potassium excretion. intersalt cooperative research group. *BMJ.* 1988;297:319-28.

122. Carvalho JJ, Baruzzi RG, Howard PF, Poulter N, Alpers MP, Franco LJ, et al. Blood pressure in four remote populations in the INTERSALT study. *Hypertension.* 1989;14:238-46.

123. Drazner MH. The progression of hypertensive heart disease. *Circulation.* 2011;123:327-34.

124. Arora P, Newton-Cheh C. Blood pressure and human genetic variation in the general population. *Curr. Opin. Cardiol.* 2010;25:229-37.

125. Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, Dehghan A, et al. Genome-wide association study of blood pressure and hypertension. *Nat. Genet.* 2009;41:677-87.

126. Newton-Cheh C, Johnson T, Gateva V, Tobin MD, Bochud M, Coin L, et al. Genome-wide association study identifies eight loci associated with blood pressure. *Nat. Genet.* 2009;41:666-76.

127. International Consortium for Blood Pressure Genome-Wide Association Studies, Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature.* 2011;478:103-9.

128. Weber C, Noels H. Atherosclerosis: Current pathogenesis and therapeutic options. *Nat. Med.* 2011;17:1410-22.

129. Roffi M, Patrono C, Collet J, Mueller C, Valgimigli M, Andreotti F, et al. 2015 ESC guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: Task force for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation of the European society of cardiology (ESC). *Eur. Heart J.* 2016;37:267-315.

130. Ibanez B, James S, Agewall S, Antunes M, Bucciarelli-Ducci C, Bueno H, et al. ESC guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation: The task force for the management of acute myocardial infarction in patients presenting with ST-segment elevation of the European society of cardiology (ESC). *Eur. Heart J.* 2017;.

131. Thygesen K, Alpert JS, Jaffe AS, Simoons ML, Chaitman BR, White HD, et al. Third universal definition of myocardial infarction. *Eur. Heart J.* 2012;33:2551-67.

132. Mollova M, Bersell K, Walsh S, Savla J, Das LT, Park SY, et al. Cardiomyocyte proliferation contributes to heart growth in young humans. *Proc. Natl. Acad. Sci. U. S. A.* 2013;110:1446-51.

133. Richardson WJ, Clarke SA, Quinn TA, Holmes JW. Physiological implications of myocardial scar structure. *Comprehensive Physiology.* 2015;5:1877-909.

134. Konstam MA, Kramer DG, Patel AR, Maron MS, Udelson JE. Left ventricular remodeling in heart failure. *JACC: Cardiovascular Imaging.* 2011;4:98-108.

135. Kannel WB, Castelli WP, Gordon T, McNamara PM. Serum cholesterol, lipoproteins, and the risk of coronary heart disease. *Ann. Intern. Med.* 1971;74:1-12.

136. Hobbs HH, Russell DW, Brown MS, Goldstein JL. The LDL receptor locus in familial hypercholesterolemia: Mutational analysis of a membrane protein. *Annu. Rev. Genet.* 1990;24:133-70.

137. Cohen JC, Boerwinkle E, Mosley Jr TH, Hobbs HH. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N. Engl. J. Med.* 2006;354:1264-72.

138. Clarke R, Frost C, Collins R, Appleby P, Peto R. Dietary lipids and blood cholesterol: Quantitative meta-analysis of metabolic ward studies. *BMJ.* 1997;314:112-7.

139. Berger S, Raman G, Vishwanathan R, Jacques PF, Johnson EJ. Dietary cholesterol and cardiovascular disease: A systematic review and meta-analysis. *Am. J. Clin. Nutr.* 2015;102:276-94.

140. McPherson R, Tybjaerg-Hansen A. Genetics of coronary artery disease. *Circ. Res.* 2016;118:564-78.

141. Helgadottir A, Thorleifsson G, Manolescu A, Gretarsdottir S, Blondal T, Jonasdottir A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science.* 2007;316:1491-3.

142. CARDIoGRAMplusC4D Consortium. A comprehensive 1000 genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* 2015;47:1121-30.

143. Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat. Genet.* 2013;45:25-33.

144. Coronary Artery Disease (C4D) Genetics Consortium. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nat. Genet.* 2011;43:339-44.

145. Schunkert H, Konig IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* 2011;43:333-8.

146. Braenne I, Civelek M, Vilne B, Di Narzo A, Johnson AD, Zhao Y, et al. Prediction of causal candidate genes in coronary artery disease loci. *Arterioscler. Thromb. Vasc. Biol.* 2015;35:2207-17.

147. Webb TR, Erdmann J, Stirrups KE, Stitziel NO, Masca NG, Jansen H, et al. Systematic evaluation of pleiotropy identifies 6 further loci associated with coronary artery disease. *J. Am. Coll. Cardiol.* 2017;69:823-36.

148. Dote K, Sato H, Tateishi H, Uchida T, Ishihara M. Myocardial stunning due to simultaneous multivessel coronary spasms: A review of 5 cases. *J. Cardiol.* 1991;21:203-14.

149. Komamura K, Fukui M, Iwasaku T, Hirotani S, Masuyama T. Takotsubo cardiomyopathy: Pathophysiology, diagnosis and treatment. *World J. Cardiol.* 2014;6:602-9.

150. Sharkey SW, Lesser JR, Maron MS, Maron BJ. Why not just call it tako-tsubo cardiomyopathy. *J. Am. Coll. Cardiol.* 2011;57:1496-7.

151. Templin C, Ghadri JR, Diekmann J, Napp LC, Bataiosu DR, Jaguszewski M, et al. Clinical features and outcomes of takotsubo (stress) cardiomyopathy. *N. Engl. J. Med.* 2015;373:929-38.

152. Dawson DK. Acute stress-induced (takotsubo) cardiomyopathy. *Heart.* 2017;104:96-102.

153. Eitel I, Von Knobelsdorff-Brenkenhoff F, Bernhardt P, Carbone I, Muellerleile K, Aldrovandi A, et al. Clinical characteristics and cardiovascular magnetic resonance findings in stress (takotsubo) cardiomyopathy. *JAMA - Journal of the American Medical Association.* 2011;306:277-86.

154. Tsuchihashi K. Transient left ventricular apical ballooning without coronary artery stenosis: A novel heart syndrome mimicking acute myocardial infarction. angina pectoris-myocardial infarction investigations in japan. *J. Am. Coll. Cardiol.* 2001;38:11-8.

155. Singh NK, Rumman S, Mikell FL, Nallamothu N, Rangaswamy C. Stress cardiomyopathy: Clinical and ventriculographic characteristics in 107 north american subjects. *Int. J. Cardiol.* 2010;141:297-303.

156. Wittstein IS. Neurohumoral features of myocardial stunning due to sudden emotional stress. *N. Engl. J. Med.* 2005;352:539-48.

157. Pison L. Apical ballooning in relatives. *Heart.* 2004;90:e67.

158. Kumar G. "Familial" apical ballooning syndrome (takotsubo cardiomyopathy). *Int. J. Cardiol.* 2010;144:444-5.

159. Prasad A, Lerman A, Rihal CS. Apical ballooning syndrome (tako-tsubo or stress cardiomyopathy): A mimic of acute myocardial infarction. *Am. Heart J.* 2008;155:408-17.

160. Spinelli L, Trimarco V, Di Marino S, Marino M, Iaccarino G, Trimarco B. L41Q polymorphism of the G protein coupled receptor kinase 5 is associated with left ventricular apical ballooning syndrome. *European journal of heart failure.* 2010;12:13-6.

161. Novo G, Giambanco S, Guglielmo M, Arvigo L, Sutera MR, Giambanco F, et al. G-protein-coupled receptor kinase 5 polymorphism and takotsubo cardiomyopathy. *J. Cardiovasc. Med. (Hagerstown).* 2015;16:639-43.

162. Goodloe AH, Evans JM, Middha S, Prasad A, Olson TM. Characterizing genetic variation of adrenergic signalling pathways in takotsubo (stress) cardiomyopathy exomes. *European journal of heart failure.* 2014;16:942-9.

163. Eitel I, Moeller C, Munz M, Stiermaier T, Meitinger T, Thiele H, et al. Genome-wide association study in takotsubo syndrome—Preliminary results and future directions. *Int. J. Cardiol.* 2017;236:335-9.

164. Cooper Jr LT. Myocarditis. *N. Engl. J. Med.* 2009;360:1526-38.

165. Marijon E, Mirabel M, Celermajer DS, Jouven X. Rheumatic heart disease. *The Lancet.* 2012;379:953-64.

166. Rassi A, Marin-Neto JA. Chagas disease. *The Lancet.* 2010;375:1388-402.

167. Cahill TJ, Ashrafian H, Watkins H. Genetic cardiomyopathies causing heart failure. *Circ. Res.* 2013;113:660-75.

168. Hershberger RE, Hedges DJ, Morales A. Dilated cardiomyopathy: The complexity of a diverse genetic architecture. *Nature reviews Cardiology.* 2013;10:531-47.

169. Herman DS, Lam L, Taylor MR, Wang L, Teekakirikul P, Christodoulou D, et al. Truncations of titin causing dilated cardiomyopathy. *N. Engl. J. Med.* 2012;366:619-28.

170. Worman HJ, Bonne G. "Laminopathies": A wide spectrum of human diseases. *Exp. Cell Res.* 2007;313:2121-33.

171. Puska P, Salonen JT, Nissinen A, Tuomilehto J, Vartiainen E, Korhonen H, et al. Change in risk factors for coronary heart disease during 10 years of a community intervention programme (North Karelia project). *Br. Med. J. (Clin. Res. Ed).* 1983;287:1840-4.

172. GTEx Consortium. Human genomics. the genotype-tissue expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science.* 2015;348:648-60.

173. Vartiainen E, Jousilahti P, Alfthan G, Sundvall J, Pietinen P, Puska P. Cardiovascular risk factor changes in Finland, 1972-1997. *Int. J. Epidemiol.* 2000;29:49-56.

174. Vartiainen E, Laatikainen T, Peltonen M, Juolevi A, Mannisto S, Sundvall J, et al. Thirty-five-year trends in cardiovascular risk factors in Finland. *Int. J. Epidemiol.* 2010;39:504-18.

175. Johansson JK, Puukka PJ, Niiranen TJ, Varis J, Peltonen M, Salomaa V, et al. Health 2000 score–development and validation of a novel cardiovascular risk score. *Ann. Med.* 2016;1-7.

176. Kristiansson K, Perola M, Tikkanen E, Kettunen J, Surakka I, Havulinna AS, et al. Genome-wide screen for metabolic syndrome susceptibility loci reveals strong lipid gene contribution but no evidence for common genetic basis for clustering of metabolic syndrome traits. *Circ. Cardiovasc. Genet.* 2012;5:242-9.

177. Eriksson JG. Early growth and coronary heart disease and type 2 diabetes: Findings from the Helsinki birth cohort study (HBCS). *Am. J. Clin. Nutr.* 2011;94:1799S-802S.

178. Jarvelin MR, Sovio U, King V, Lauren L, Xu B, McCarthy MI, et al. Early life factors and blood pressure at age 31 years in the 1966 northern Finland birth cohort. *Hypertension.* 2004;44:838-46.

179. Vaara S, Nieminen MS, Lokki M, Perola M, Pussinen PJ, Allonen J, et al. Cohort profile: The Corogene study. *Int. J. Epidemiol.* 2012;41:1265-71.

180. Nikus KC, Eskola MJ, Virtanen VK, Harju J, Huhtala H, Mikkelsson J, et al. Mortality of patients with acute coronary syndromes still remains high: A follow-up study of 1188 consecutive patients admitted to a university hospital. *Ann. Med.* 2007;39:63-71.

181. Parkkonen O. Takotsubo cardiomyopathy—Differentiation from acute coronary syndrome by electrocardiography and biochemistry [dissertation]. University of Helsinki; 2017.

182. Paavonen EJ, Saarenpää-Heikkilä O, Pölkki P, Kylliäinen A, Porkka-Heiskanen T, Paunio T. Maternal and paternal sleep during pregnancy in the child-sleep birth cohort. *Sleep Med.* 2017;29:47-56.

183. Bumgarner R. Overview of DNA microarrays: Types, applications, and their future. *Current protocols in molecular biology.* 2013;22.1. 1,22.1. 11.

184. Pajunen P, Koukkunen H, Ketonen M, Jerkkola T, Immonen-Räihä P, Kärjä-Koskenkari P, et al. The validity of the Finnish hospital discharge register and causes of death register data on coronary heart disease. *European Journal of Cardiovascular Prevention & Rehabilitation.* 2005;12:132-7.

185. Blankenberg S, Zeller T, Saarela O, Havulinna AS, Kee F, Tunstall-Pedoe H, et al. Contribution of 30 biomarkers to 10-year cardiovascular risk estimation in 2 population cohorts: The MONICA, risk, genetics, archiving, and monograph (MORGAM) biomarker project. *Circulation.* 2010;121:2388-97.

186. IFCC-international federation of clinical chemistry and laboratory medicine: Analytical characteristics of commercial MR-proANP, BNP and NT-proBNP assays as per the manufacturer. Available from: http://www.ifcc.org/media/276711/IFCC%20NP%20Assay%20Table%20November%202014.pdf.

187. Inouye M, Kettunen J, Soininen P, Silander K, Ripatti S, Kumpula LS, et al. Metabonomic, transcriptomic, and genomic variation of a population cohort. *Molecular Systems Biology.* 2010;6:441.

188. Inouye M, Silander K, Hamalainen E, Salomaa V, Harald K, Jousilahti P, et al. An immune response network associated with blood lipid levels. *PLoS genetics.* 2010;6:e1001113.

189. Aguet F, Ardlie KG, Cummings BB, Gelfand ET, Getz G, Hadley K, et al. Genetic effects on gene expression across human tissues. *Nature.* 2017;550:204-13.

190. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS genetics.* 2009;5:e1000529.

191. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: Using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* 2010;34:816-34.

192. Delaneau O, Marchini J, Zagury J. A linear complexity phasing method for thousands of genomes. *Nature Methods.* 2012;9:179-81.

193. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature.* 2007;449:851-61.

194. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature.* 2010;467:1061-73.

195. International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature.* 2010;467:52-8.

196. Surakka I, Kristiansson K, Anttila V, Inouye M, Barnes C, Moutsianas L, et al. Founder population-specific HapMap panel increases power in GWA studies through improved imputation accuracy and CNV tagging. *Genome Res.* 2010;20:1344-51.

197. 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature.* 2012;491:56-65.

198. Surakka I, Sarin A, Ruotsalainen SE, Durbin R, Salomaa V, Daly M, et al. The rate of false polymorphisms introduced when imputing genotypes from global imputation panels. *bioRxiv.* 2016;.

199. Buuren S, Groothuis-Oudshoorn K. Mice: Multivariate imputation by chained equations in R. *Journal of statistical software.* 2011;45:1-68.

200. Schafer JL. Multiple imputation: A primer. *Stat. Methods Med. Res.* 1999;8:3-15.

201. Beasley TM, Erickson S, Allison DB. Rank-based inverse normal transformations are increasingly used, but are they merited? *Behav. Genet.* 2009;39:580.

202. Aulchenko YS, Struchalin MV, van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics.* 2010;11:134.

203. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* 2007;39:906-13.

204. Liu JZ, Tozzi F, Waterworth DM, Pillai SG, Muglia P, Middleton L, et al. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat. Genet.* 2010;42:436-40.

205. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 2007;81:559-75.

206. Magi R, Morris AP. GWAMA: Software for genome-wide association meta-analysis. *BMC Bioinformatics.* 2010;11:288,2105-11-288.

207. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2013.

208. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw.* 2010;36:1-48.

209. Sabourin J, Nobel AB, Valdar W. Fine-Mapping additive and dominant SNP effects using Group-LASSO and fractional resample model averaging. *Genet. Epidemiol.* 2015;39:77-88.

210. Therneau TM. A package for survival analysis in S. 2015;.

211. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex trait analysis. *The American Journal of Human Genetics.* 2011;88:76-82.

212. Gaal EI, Salo P, Kristiansson K, Rehnstrom K, Kettunen J, Sarin AP, et al. Intracranial aneurysm risk locus 5q23.2 is associated with elevated systolic blood pressure. *PLoS Genet.* 2012;8:e1002563.

213. Yasuno K, Bakircioglu M, Low SK, Bilguvar K, Gaal E, Ruigrok YM, et al. Common variant near the endothelin receptor type A (EDNRA) gene is associated with intracranial aneurysm risk. *Proc. Natl. Acad. Sci. U. S. A.* 2011;108:19707-12.

214. Davis CA, Haberland M, Arnold MA, Sutherland LB, McDonald OG, Richardson JA, et al. PRISM/PRDM6, a transcriptional repressor that promotes the proliferative gene program in smooth muscle cells. *Mol. Cell. Biol.* 2006;26:2626-36.

215. Zhao XQ, Theroux P, Snapinn SM, Sax FL. Intracoronary thrombus and platelet glycoprotein IIb/IIIa receptor blockade with tirofiban in unstable angina or non-Q-wave myocardial infarction. angiographic results from the PRISM-PLUS trial (platelet receptor inhibition for ischemic syndrome management in patients limited by unstable signs and symptoms). PRISM-PLUS investigators. *Circulation.* 1999;100:1609-15.

216. Stone GW, Cox D, Garcia E, Brodie BR, Morice M-, Griffin J, et al. Normal flow (TIMI-3) before mechanical reperfusion therapy is an independent determinant of survival in acute myocardial infarction: Analysis from the primary angioplasty in myocardial infarction trials. *Circulation.* 2001;104:636.

217. Salo PP, Vaara S, Kettunen J, Pirinen M, Sarin AP, Huikuri H, et al. Genetic variants on chromosome 1p13.3 are associated with non-ST elevation myocardial infarction and the expression of DRAM2 in the Finnish population. *PLoS One.* 2015;10:e0140576.

218. Reilly MP, Mingyao L, Jing H, Ferguson JF, Stylianou IM, Mehta NN, et al. Identification of ADAMTS7 as a novel locus for coronary atherosclerosis and association of ABO with myocardial infarction in the presence of coronary atherosclerosis: Two genome-wide association studies. *Lancet.* 2011;377:383-92.

219. O'Connor CM, Starling RC, Hernandez AF, Armstrong PW, Dickstein K, Hasselblad V, et al. Effect of nesiritide in patients with acute decompensated heart failure. *N. Engl. J. Med.* 2011;365:32-43.

220. Pfister R, Sharp S, Luben R, Welsh P, Barroso I, Salomaa V, et al. Mendelian randomization study of B-type natriuretic peptide and type 2 diabetes: Evidence of causal association from population studies. *PLoS Med.* 2011;8:e1001112.

221. Lek M, Karczewski K, Minikel E, Samocha K, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *BioRxiv.* 2016;030338.

222. Molkentin JD, Lu J, Antos CL, Markham B, Richardson J, Robbins J, et al. A calcineurin-dependent transcriptional pathway for cardiac hypertrophy. *Cell.* 1998;93:215-28.

223. Eitel I, Moeller C, Munz M, Stiermaier T, Meitinger T, Thiele H, et al. Genome-wide association study in takotsubo syndrome—Preliminary results and future directions. *Int. J. Cardiol.* 2017;236:335-9.

224. Figtree GA, Bagnall RD, Abdulla I, Buchholz S, Galougahi KK, Yan W, et al. No association of G-protein-coupled receptor kinase 5 or beta-adrenergic receptor polymorphisms with takotsubo cardiomyopathy in a large australian cohort. *Eur. J. Heart Fail.* 2013;15:730-3.

225. Goodloe AH, Evans JM, Middha S, Prasad A, Olson TM. Characterizing genetic variation of adrenergic signalling pathways in takotsubo (stress) cardiomyopathy exomes. *Eur. J. Heart Fail.* 2014;16:942-9.

226. Novo G, Giambanco S, Guglielmo M, Arvigo L, Sutera MR, Giambanco F, et al. G-protein-coupled receptor kinase 5 polymorphism and takotsubo cardiomyopathy. *J. Cardiovasc. Med. (Hagerstown).* 2015;16:639-43.

227. Spinelli L, Trimarco V, Di Marino S, Marino M, Iaccarino G, Trimarco B. L41Q polymorphism of the G protein coupled receptor kinase 5 is associated with left ventricular apical ballooning syndrome. *Eur. J. Heart Fail.* 2010;12:13-6.

228. Sharkey SW, Lips DL, Pink VR, Maron BJ. Daughter-mother tako-tsubo cardiomyopathy. *Am. J. Cardiol.* 2013;112:137-8.

229. Kuznetsova T, Citterio L, Zagato L, Delli Carpini S, Thijs L, Casamassima N, et al. Left ventricular radial function associated with genetic variation in the cGMP-dependent protein kinase. *Hypertension.* 2013;62:1034-9.

230. Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, Buros J, et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat. Genet.* 2011;43:436-41.

231. Kapoor A, Jiang Q, Chatterjee S, Chakraborty P, Sosa MX, Berrios C, et al. Population variation in total genetic risk of hirschsprung disease from common RET, SEMA3 and NRG1 susceptibility polymorphisms. *Hum. Mol. Genet.* 2015;24:2997-3003.

232. Ehret GB, Ferreira T, Chasman DI, Jackson AU, Schmidt EM, Johnson T, et al. The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. *Nat. Genet.* 2016;48:1171-84.

233. Warren HR, Evangelou E, Cabrera CP, Gao H, Ren M, Mifsud B, et al. Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nat. Genet.* 2017;49:403-15.

234. Park S, Kim K, Lee E, Kim B, Lee TJ, Seo T, et al. Reduced expression of DRAM2/TMEM77 in tumor cells interferes with cell death. *Biochem. Biophys. Res. Commun.* 2009;390:1340-4.

235. Tata JR. One hundred years of hormones. *EMBO Rep.* 2005;6:490-6.

236. John SW, Krege JH, Oliver PM, Hagaman JR, Hodgin JB, Pang SC, et al. Genetic decreases in atrial natriuretic peptide and salt-sensitive hypertension. *Science.* 1995;267:679-81.

237. Steinhelper ME, Cochrane KL, Field LJ. Hypotension in transgenic mice expressing atrial natriuretic factor fusion genes. *Hypertension.* 1990;16:301-7.

238. Ogawa Y, Itoh H, Tamura N, Suga S, Yoshimasa T, Uehira M, et al. Molecular cloning of the complementary DNA and gene that encode mouse brain natriuretic peptide and generation of transgenic mice that overexpress the brain natriuretic peptide gene. *J. Clin. Invest.* 1994;93:1911-21.

239. Tamura N, Ogawa Y, Chusho H, Nakamura K, Nakao K, Suda M, et al. Cardiac fibrosis in mice lacking brain natriuretic peptide. *Proc. Natl. Acad. Sci. U. S. A.* 2000;97:4239-44.

240. Crozier I, Ikram H, Gomez H, Nicholls MG, Espiner E, Warner NJ. Haemodynamic effects of atrial peptide infusion in heart failure. *The Lancet.* 1986;328:1242-5.

241. Hynynen M, Kupari M, Salmenperä M, Tikkanen I, Heinonen J, Fyhrquist F, et al. Hemodynamic effects of [alpha]-human atrial natriuretic peptide in healthy volunteers. *J. Cardiovasc. Pharmacol.* 1988;11:711-5.

242. Jensen KT, Carstens J, Pedersen EB. Effect of BNP on renal hemodynamics, tubular function and vasoactive hormones in humans. *Am. J. Physiol.* 1998;274:F63-72.

243. Suga S, Nakao K, Hosoda K, Mukoyama M, Ogawa Y, Shirakami G, et al. Receptor selectivity of natriuretic peptide family, atrial natriuretic peptide, brain natriuretic peptide, and C-type natriuretic peptide. *Endocrinology.* 1992;130:229-39.

244. Pidgeon GB, Richards AM, Nicholls MG, Espiner EA, Yandle TG, Frampton C. Differing metabolism and bioactivity of atrial and brain natriuretic peptides in essential hypertension. *Hypertension.* 1996;27:906-13.

245. Asferg CL, Andersen UB, Linneberg A, Hedley PL, Christiansen M, Goetze JP, et al. Serum proatrial natriuretic peptide does not increase with higher systolic blood pressure in obese men. *Heart.* 2017;103:154-8.

246. Coffman TM. Under pressure: The search for the essential mechanisms of hypertension. *Nat. Med.* 2011;17:1402-9.

247. Padmanabhan S, Newton-Cheh C, Dominiczak AF. Genetic basis of blood pressure and hypertension. *Trends in Genetics.* 2012;28:397-408.

248. Semenov AG, Tamm NN, Seferian KR, Postnikov AB, Karpova NS, Serebryanaya DV, et al. Processing of pro-B-type natriuretic peptide: Furin and corin as candidate convertases. *Clin. Chem.* 2010;56:1166-76.

249. Gee CE, Peterlik D, Neuhauser C, Bouhelal R, Kaupmann K, Laue G, et al. Blocking metabotropic glutamate receptor subtype 7 (mGlu7) via the venus flytrap domain (VFTD) inhibits amygdala plasticity, stress, and anxiety-related behavior. *J. Biol. Chem.* 2014;289:10975-87.

250. Masugi M, Yokoi M, Shigemoto R, Muguruma K, Watanabe Y, Sansig G, et al. Metabotropic glutamate receptor subtype 7 ablation causes deficit in fear response and conditioned taste aversion. *J. Neurosci.* 1999;19:955-63.

251. Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: From polygenic to omnigenic. *Cell.* 2017;169:1177-86.